IN-17
146080
P-61

NASA
Technical
Paper
3180

January 1993

Optical Communications and a Comparison of Optical Technologies for a High Data Rate Return Link From Mars

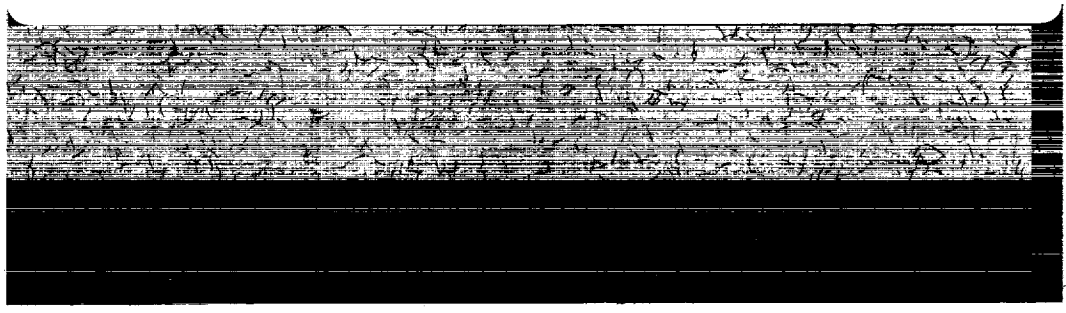
Rodney L. Spence

(NASA-TP-3180) OPTICAL COMMUNICATIONS AND A COMPARISON OF OPTICAL TECHNOLOGIES FOR A HIGH DATA RATE RETURN LINK FROM MARS (NASA) 61 p

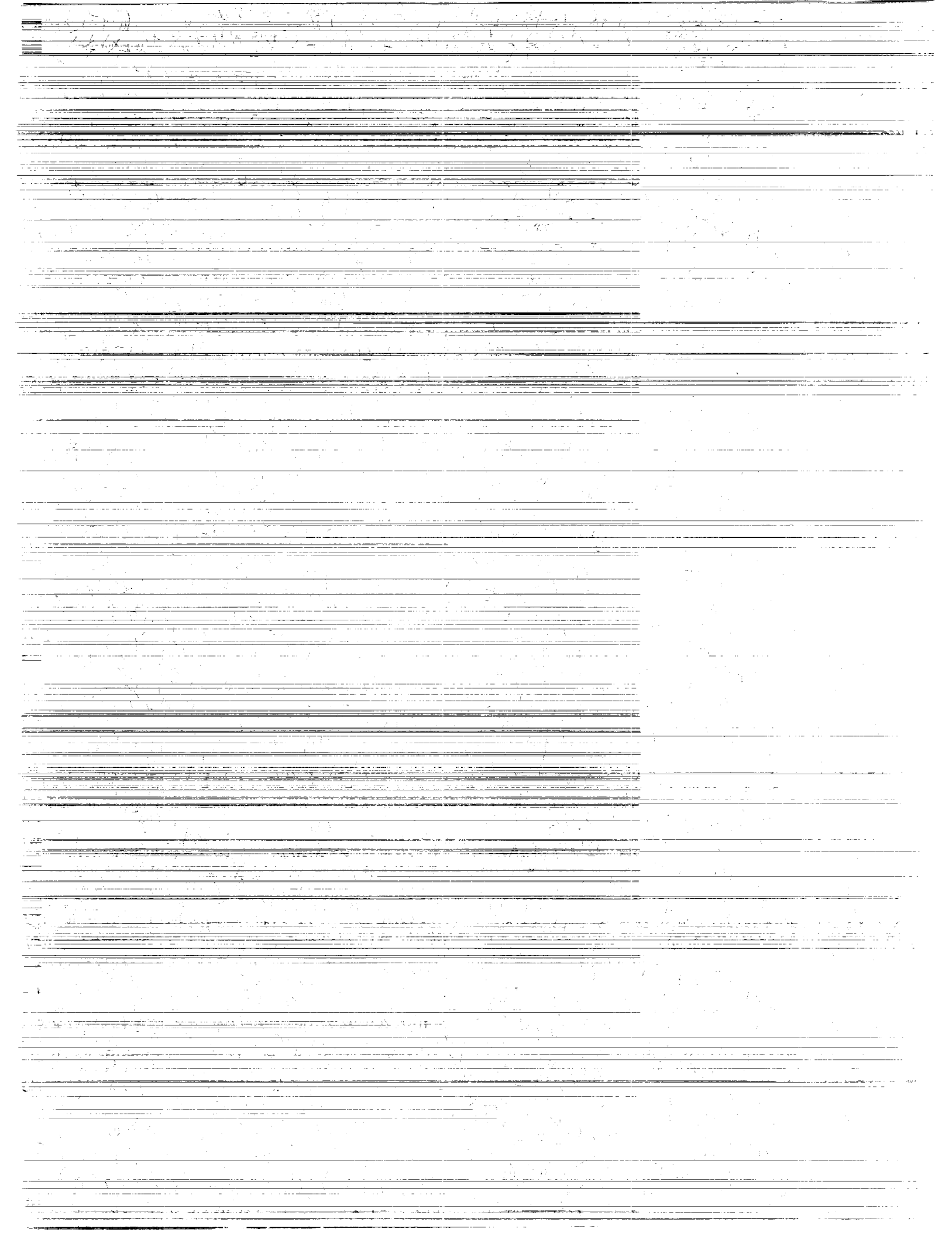
N93-18854

Unclas

H1/17 0146080



NASA



NASA Technical Paper 3180

1993

Optical Communications and a Comparison of Optical Technologies for a High Data Rate Return Link From Mars

Rodney L. Spence
Lewis Research Center
Cleveland, Ohio

NASA
National Aeronautics and Space Administration
Office of Management
Scientific and Technical Information Program

Table of Contents

PRECEDING PAGE BLANK NOT FILMED

Summary

The important principles of direct- and heterodyne-detection optical free-space communications are reviewed. Signal-to-noise-ratio (SNR) and bit-error-rate (BER) expressions are derived for both the direct-detection and heterodyne-detection optical receivers. For the heterodyne system, performance degradation resulting from received-signal and local-oscillator-beam misalignment and laser phase noise is analyzed. Determination of interfering background power from local- and extended-background sources is discussed. The BER performance of direct- and heterodyne-detection optical links in the presence of Rayleigh-distributed random pointing and tracking errors is described. Finally, several optical systems employing Nd:YAG (neodymium yttrium aluminum garnet), GaAs (gallium arsenide), and CO_2 (carbon dioxide) laser sources are evaluated and compared to assess their feasibility in providing high-data-rate (10- to 1000-Mbps) Mars-to-Earth communications.

It is shown that the root mean square (rms) pointing and tracking accuracy is a critical factor in defining the system transmitting laser-power requirements and telescope size and that, for a given rms error, there is an optimum telescope aperture size that minimizes the required power. The results of the analysis conducted indicate that, barring the achievement of extremely small rms pointing and tracking errors (<0.2 μrad), the two most promising types of optical systems are those that use an Nd:YAG laser ($\lambda = 1.064$ μm) and high-order pulse position modulation (PPM) and direct detection, and those that use a CO_2 laser ($\lambda = 10.6$ μm) and phase shift keying (PSK) homodyne modulation and coherent detection. For example, for a PPM order of $M = 64$ and an rms pointing accuracy of 0.4 μrad, an Nd:YAG system can be used to implement a 100-Mbps Mars link with a 40-cm transmitting telescope, a 20-W laser, and a 10-m receiving photon bucket. Under the same conditions, a CO_2 system would require 3-m transmitting and receiving telescopes and a 32-W laser to implement such a link. Other types of optical systems, such as semiconductor laser systems, are impractical in the presence of large rms pointing errors because of the high power requirements of the 100-Mbps Mars link, even when optimal-size telescopes are used.

Introduction

This report has two objectives: to present the important principles and equations necessary for evaluating the performance of direct-detection and heterodyne-detection optical-free-space communication links (sections 1 to 6) and to draw on this material to assess the feasibility of supporting a high-data-rate (10- to-1000 Mbps) Mars-to-Earth communication link with each of several different optical systems (section 7). Future manned Mars missions planned as part of the Space Exploration Initiative (SEI) will require such high transmission rates to carry high-resolution video images and scientific-instrument data. Optical systems appear to offer size, weight, and power advantages over equivalent radiofrequency (RF) systems in providing such high transmission rates.

In sections 1 and 2, the quantum-limited signal-to-noise ratios (SNR's) are derived for the direct-detection receiver and heterodyne-detection receiver, respectively.

Sections 3 and 4 examine the development of bit-error-rate (BER) expressions for the two detection types. For the direct-detection system, BER is computed assuming M-ary pulse position modulation (PPM) and an avalanche photodiode (APD) based receiver. For the heterodyne system, BER is presented for a number of modulation types. The BER expressions for heterodyne detection are shown to easily follow from the corresponding RF expressions under a strong, local oscillator (LO) laser condition. Also discussed are losses unique to heterodyne detection, such as laser phase noise and LO beam and receiver signal-beam misalignment—an important phenomenon because heterodyne systems retain their sensitivity advantage over direct-detection systems only when these two beams are precisely aligned and are mutually spatially coherent.

Section 5 presents equations for determining the level of interfering background power from local and extended background sources modeled as blackbody radiators. It is shown that the heterodyne receiver is relatively immune to thermal background interference even when the Sun is in the receiver's field of view.

Because optical beam widths are extremely narrow, spatial pointing and tracking errors can have a marked effect on link performance. Consequently, section 6 considers the BER performance of both direct- and heterodyne-detection systems in the presence of random pointing and tracking errors. The pointing error along each of the two (orthogonal) telescope gimbal axes is modeled as a zero-mean gaussian random variable with the root mean square (rms) value σ (i.e., the radial pointing error is Rayleigh distributed). Under these conditions, it is shown that, for a given rms error, an optimal transmitter-aperture size exists for the direct-detection system that minimizes the required transmitter power for a given BER performance. Similarly, it

is shown that optimal transmitter and receiver aperture sizes exist for the heterodyne system that minimize the required transmission power. For a direct-detection system, the minimal power is shown to be proportional to the square of the rms error, whereas, for a heterodyne detection system, the power is shown to be proportional to the fourth power of the rms error. Consequently, it is demonstrated that the heterodyne receiver performs better at a given wavelength than does the direct-detection receiver only if the rms pointing and tracking error can be kept very small. As such, this section shows that, in an environment characterized by large rms error, the direct-detection system is preferable despite its lower receiver sensitivity.

Section 6 also discusses the difference between random pointing error and static pointing error and shows that an analysis which assumes static pointing-error conditions may yield significantly more optimistic results than an analysis that realistically models the pointing error as a random variable.

Section 7 analyzes the suitability of seven optical systems for use as high-data-rate Mars-to-Earth communication links. Optical implementations based on Nd:YAG, GaAs, and CO_2 laser sources and employing both direct detection and heterodyne detection are considered. Through the modeling of pointing and tracking errors as Rayleigh random variables, optimal aperture sizes and minimal transmission power requirements are determined for each optical system as a function of the link data rate and the rms error. An optical system with a Nd:YAG laser and high-order PPM and direct detection, and another with a CO_2 laser that uses homodyne phase shift keying (PSK) modulation and coherent detection emerge as the two most promising candidates for a high-data-rate Mars-to-Earth communication link. Because of their limited power output, other optical systems, such as those using a GaAs diode laser, appear practical only for lower data rate links (~10 Mbps).

A comprehensive list of symbols and acronyms is provided in appendix A.

1.0 Signal-to-Noise Ratio for the Direct-Detection Optical Receiver

In this section, the quantum-limited SNR of a direct-detection optical receiver is derived. The SNR is the appropriate measure of performance for an analog communication system since the objective at the receiver is to reproduce the transmitted waveform shape in the presence of noise. In contrast to the case for an RF system, it is shown that, because of signal-generated shot noise, the SNR is finite even when all thermal noise and background noise is ignored. The use of an APD to overcome detector thermal noise also is considered here.

Figure 1.1 depicts a block diagram of a typical direct-detection optical receiver. After being collected by the receiving optics, the optical signal passes through a field-stop iris, collimating lens, and optical bandpass filter, and is focused

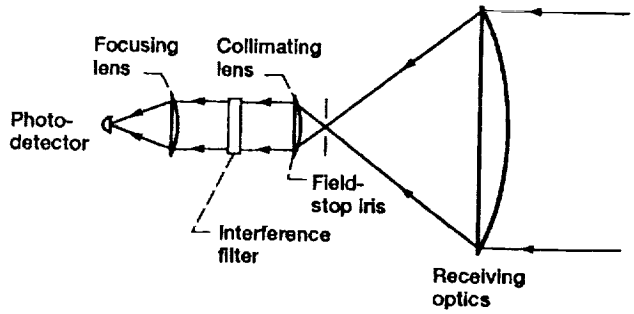


Figure 1.1.—Block diagram of a direct-detection receiver.

on the active region of a photodetector. The field-stop iris controls the receiver field of view (FOV) while the optical bandpass filter is used to reject unwanted background radiation. The detector typically is a semiconductor PIN (P-type, intrinsic, N-type) photodiode or an APD. The photodetector—acting as a transducer—generates an output current that is proportional to the instantaneous intensity of the incident optical signal. Since the detector responds only to fluctuations in field intensity and not in frequency or phase, direct detection also is called noncoherent detection. The received electric field of an unmodulated carrier may be expressed as

$$E(t) = A \cos (2\pi f_s t + \theta) \quad \text{V/m} \tag{1.1}$$

where

A electric field amplitude, V/m

f_s optical frequency, Hz

θ optical carrier phase

The corresponding field intensity is then defined as

$$I = \frac{1}{Z_o}\overline{E(t)^2} = \frac{A^2}{2Z_o} \quad \text{W/m}^2 \tag{1.2}$$

where the time average

$$\overline{E(t)^2} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} E(t)^2 dt$$

and the free space impedance $Z_o \approx 377\Omega$.

Note that the electric field's intensity is proportional to the square of its amplitude. If the information signal is $m(t)$, then the intensity of an intensity modulated (IM) carrier is given by

$$I(t) = I_S[1 + m(t)] \tag{1.3}$$

where I_S is the average received signal intensity. The degree of intensity modulation is described by the intensity modulation index

$$M_{IM} = \frac{[I(t)]_{MAX} - I_S}{I_S} \tag{1.4}$$

When $|m(t)| \leq 1$, the modulation index is equal to 1 and the carrier is said to be 100-percent intensity modulated. To prevent overmodulation, we assume $m(t)$ is normalized so that $|m(t)| \leq 1$. In this case, the average power of $m(t)$ is

$$P_m = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} m(t)^2 dt \leq 1 \tag{1.5}$$

Multiplying the intensity by the receiver collecting area A_r results in an instantaneous received signal power of

$$P_S(t) = A_r I(t) = A_r I_S + A_r I_S m(t) = P_S + P_S m(t) \tag{1.6}$$

where P_S is the average received signal power. The average received power is related to the average transmitted power through the range equation

$$P_S = P_t \eta_t \left(\frac{\pi d_t}{\lambda}\right)^2 \left(\frac{\lambda}{4\pi z}\right)^2 L_t(\theta_t, d_t) \eta_r \left(\frac{\pi d_r}{\lambda}\right)^2 L_o \tag{1.7}$$

where

P_t	average transmitter laser power, W
η_t, η_r	transmitter and receiver optics efficiencies
d_t, d_r	transmitter and receiver aperture diameters, m
λ	optical wavelength, m
z	link distance between transmitter and receiver, m
L_o	other losses (e.g., obscuration, implementation, filter)
$L_t(\theta_t, d_t)$	transmitter pointing loss (see section 6)

The component of photodetector output current resulting from the desired signal is

$$i_d(t) = \frac{\eta}{h\nu} e P_S + \frac{\eta}{h\nu} e P_S m(t) \tag{1.8}$$

where

η	photodetector quantum efficiency, (number of emitted photoelectrons/number of incident photons)
h	Planck constant, 6.624×10^{-34} J/sec
ν	optical frequency, Hz
$h\nu$	photon energy, J/photon
e	electron charge, 1.6×10^{-19} C

Hereafter, we will let $\alpha = \eta/h\nu$ be the conversion factor that converts average received power (in watts) into the average rate of photoelectron emission from the detector (in photoelectrons per second). If we assume that the post-detection processor consists of an electrical low-pass filter with a unity-transfer function over the information-signal bandwidth B_m, then, after removal of the direct current (dc) term in equation (1.8), the signal current at the output of the filter is

$$i_S(t) = \alpha e P_S m(t) \tag{1.9}$$

The corresponding average signal power is then given by the mean-square value, or

$$S = \overline{i_S(t)^2} = \alpha^2 e^2 P_S^2 P_m \tag{1.10}$$

In addition to the signal component, the output-filter current has components caused by detector shot noise and circuit thermal noise; shot noise resulting from the signal, background radiation, and detector dark current are generated within the detector and combine with thermal noise at the detector output. (Dark current is the random emission of electrons from the detector surface resulting from thermal excitation and is so named because it exists even when no radiation is incident on the detector).

Shot noise is caused by the random emission of electrons from the detector surface. When light of constant power P is incident on the detector, the number of electrons emitted per unit time is not constant, but instead follows a Poisson distribution. Since each electron may be considered to contribute a pulse to the total current, the emission of the electrons at random produces random fluctuations around a mean value. Hence, the output current from the detector consists of an average current, $\bar{I} = \alpha e P$, with random noise added. The one-sided power spectral density (PSD) of the shot noise is flat with the value $2e\bar{I}$ (ref. 1). Consequently, shot noise cannot be overcome simply by increasing the incident power P because shot noise also increases with power. Furthermore, shot noise always will be present—even in the absence of thermal and background noise—since it is generated during detection of the desired signal. The existence of shot noise also derives from the discrete nature of the charge carrier; for example, if the charge of an individual electron were to approach zero, the shot noise also would tend to zero.

The thermal, or Johnson, noise is the same as that present in RF receivers, and it results from the thermal fluctuation of electrons that follows the photodetector in the resistive elements of the electronic circuitry. If the detector load resistance is R_L and the effective noise temperature of the detector preamplifier is T, then the one-sided power spectral density of the thermal noise is $4kT/R_L$.

After passing through the low-pass filter of bandwidth B_m, the total shot-noise power resulting from signal, background radiation, and dark current is

$$2eI_S B_m + 2eI_B B_m + 2eI_D B_m$$
$$= 2e^2 \alpha P_S B_m + 2e^2 \alpha P_B B_m + 2eI_D B_m \tag{1.11}$$

where

I_S average signal current, $\alpha e P_S$

I_B average background current, $\alpha e P_B$

I_D average dark current

P_B average received background power

(The calculation of the received background radiation power is described in section 5). The thermal noise power at the filter output is

$$\frac{4kTB_m}{R_L} \tag{1.12}$$

Using equations (1.10) to (1.12) yields an SNR at the receiver output in terms of the average received signal and background power:

$$\frac{S}{N} = \frac{\alpha^2 e^2 P_S^2 P_m}{2e^2 \alpha B_m (P_S + P_B) + 2eI_D B_m + \dfrac{4kTB_m}{R_L}} \tag{1.13}$$

If the received signal and background power are large enough that the thermal noise can be neglected, the receiver is termed shot-noise limited and the SNR becomes

$$\frac{S}{N} = \frac{\alpha^2 e^2 P_S^2 P_m}{2e^2 \alpha B_m (P_S + P_B) + 2eI_D B_m} \tag{1.14}$$

Finally, if $P_S \gg P_B$, the dark-current shot noise is negligible and the modulating signal $m(t)$ has a maximum power of $P_m = 1$, then the receiver is termed signal shot-noise limited, and

$$\frac{S}{N} = \frac{\alpha^2 e^2 P_S^2}{2e^2 \alpha B_m P_S} = \frac{\alpha P_S}{2B_m} = \frac{\eta P_S}{2h\nu B_m} \tag{1.15}$$

For direct detection, this SNR also is called the quantum-limited SNR because the denominator can be thought of as the noise power occurring in a bandwidth B_m as a result of an effective quantum noise whose two-sided spectral-density level is equal to $h\nu$. Note that even under ideal conditions, wherein background noise, thermal noise, and dark current noise are wholly absent, the SNR for direct detection is not infinite as it would be in an RF system, but it is limited by the shot noise generated by the desired signal. The shot noise is not additive, as is thermal noise in an RF system; rather, it is an inescapable consequence of photodetection. The presence of signal-generated shot noise in an optical receiver is the key factor that distinguishes such a receiver from an RF receiver. Since shot noise cannot be eliminated from the optical receiver, the goal in optical detection is to drive the receiver into a shot-noise-limited operating mode, thus bringing the SNR as close to the quantum limit as possible.

One way of achieving quantum-noise-limited performance is simply to provide enough transmission power for the received signal power P_S to overwhelm the background and thermal noise. However, because of the limitations of transmitter laser power and the large link distances, this brute-force technique is not practical for attaining the shot-noise-limited condition.

A preferred method of attaining the shot-noise-limited condition is to employ photomultiplication in the photodetection process by using an APD. APD's operate under high reverse-bias voltages. Consequently, when primary carriers (electron-hole pairs) are generated by incident photons, the primary carriers gain sufficient energy to ionize other electron-hole pairs upon colliding with them. These newly created carriers are then accelerated and generate still more carriers through impact ionization. As a result of these processes, the APD provides an internal gain that enables the multiplication of the primary photodetector current prior to its encounter with the thermal noise in the receiver circuitry. Note that, since background-noise photons are photomultiplied along with desired-signal photons, APD's are effective only in overcoming thermal noise, not in overcoming background noise. (In a direct-detection system, background noise is controlled by high signal-power transmission and by the use of a narrow-band optical filter at the receiver.)

Typical APD gains range from 100 to 300, but because the avalanche mechanism is a statistical process, these are average rather than constant gain values. In other words, the APD's gain is itself a random variable, independent from one carrier pair to the next. Although the gain agrees closely with a McIntyre probability distribution (refs. 2 and 3), it is normally approximated with gaussian statistics for reasons of mathematical tractability. The randomness in the carrier multiplication is called avalanche gain noise. The effect of the random gain G is to increase the desired-signal power by the squared mean of the gain $(\bar{G})^2$ and the shot-noise power by the mean squared gain $\overline{G^2}$ (ref. 4). Therefore, from equation (1.13), we obtain the direct detection SNR with photomultiplication

$$\frac{S}{N} = \frac{(\bar{G})^2 \alpha^2 e^2 P_S^2 P_m}{2\overline{G^2} e^2 \alpha B_m (P_S + P_B) + 2e\overline{G^2} I_D B_m + \dfrac{4kTB_m}{R_L}} \tag{1.16}$$

which, if we neglect dark current, can be expressed in the form

$$\frac{S}{N} = \frac{\alpha P_S P_m}{2\dfrac{\overline{G^2}}{(\bar{G})^2} B_m \left(1 + \dfrac{P_B}{P_S}\right) + \dfrac{4kTB_m}{(\bar{G})^2 e^2 R_L \alpha P_S}} \tag{1.17}$$

From this equation we see that the thermal-noise term can be neglected if the average gain of the APD is large enough. Since the two-sided thermal-noise spectral density is $N_T = 2kT/R_L$ and the two-sided signal shot-noise density is $N_S = \alpha e^2 P_S$, we must obtain an average gain such that

$$(\bar{G})^2 e^2 \alpha P_S \gg \frac{2kT}{R_L} \quad \text{or} \quad \bar{G} \gg \sqrt{\frac{2kT/R_L}{e^2 \alpha P_S}} = \sqrt{\frac{N_T}{N_S}} \tag{1.18}$$

Under this condition, the shot-noise-limited SNR is

$$\frac{S}{N} = \frac{(\bar{G})^2}{\overline{G^2}} \left[\frac{\alpha P_S P_m}{2B_m(1 + P_B/P_S)} \right] \tag{1.19}$$

As such, if the average received signal power is much greater than the background radiation and the modulating signal $m(t)$ has its maximum power $P_m = 1$, the signal shot-noise-limited SNR becomes

$$\frac{S}{N} = \frac{(\bar{G})^2}{\overline{G^2}} \left(\frac{\alpha P_S}{2B_m} \right) = \frac{(\bar{G})^2}{\overline{G^2}} \left(\frac{\eta P_S}{2h\nu B_m} \right) \tag{1.20}$$

This is simply the SNR of equation (1.15) multiplied by the additional $(\bar{G})^2/\overline{G^2}$ factor resulting from photomultiplication. The reciprocal of this factor is defined as the APD excess-noise factor and can be written as

$$F = \frac{\overline{G^2}}{(\bar{G})^2} = \frac{(\bar{G})^2 + \sigma_G^2}{(\bar{G})^2} = 1 + \frac{\sigma_G^2}{(\bar{G})^2} \tag{1.21}$$

which follows from the definition of variance. The term σ_G^2 is the variance of G and is a measure of the randomness of the APD gain—the larger the gain spread of the device, the greater the value of F. Thus, the penalty for using an APD to overcome thermal noise and to attain shot-noise-limited behavior is the introduction of this multiplication noise. Note that an ideal APD would exhibit constant gain and zero variance. Such conditions result in $F = 1$ and allow achievement of the quantum-limited SNR (provided that the signal power is still large enough that background and dark-current noise can be ignored). Typical silicon APD's have excess noise factors in the range 2 to 5.

2.0 Optical Heterodyne Detection

The use of optical heterodyning provides an alternate means of achieving near quantum-limited behavior of the optical receiver. The basic structure of a heterodyne receiver is shown in figure 2.1. Heterodyne detection is based on the interference that results when two mutually coherent (defined in section 2.1) electromagnetic waves overlap in space. In the heterodyne receiver, the optical, or electric, field produced by a laser LO is optically mixed with the received-signal optical field prior to photodetection. The detector then responds to the combined field as if it were a single received field. If the LO frequency differs from the received optical carrier frequency, the detection operation is termed heterodyning. If the two frequencies are the same, the detection operation is termed homodyning. If the optical and received signal fields are spatially aligned by means of the partially transmitting mirror then, as shown below, the resulting optical interference generates an intensity term that is linearly proportional to the amplitude, frequency, and phase of the received optical carrier. As such, each of these carrier parameters may be modulated by the information signal, enabling amplitude (AM), frequency (FM), or phase (PM) modulation—in contrast to the case in direct detection. In addition, if the LO laser power is made sufficiently large, a near quantum-limited SNR can be obtained, as shown in section 2.1. It should be emphasized that the heterodyning and homodyning operations are not performed by the detector but rather are a consequence of the signal and LO optical fields combining at the photodetector surface. As in direct detection, in homodyne and heterodyne detection, the detector itself can respond only to variations in field intensity.

High heterodyning efficiency requires that the received signal field and LO fields be precisely aligned along the same

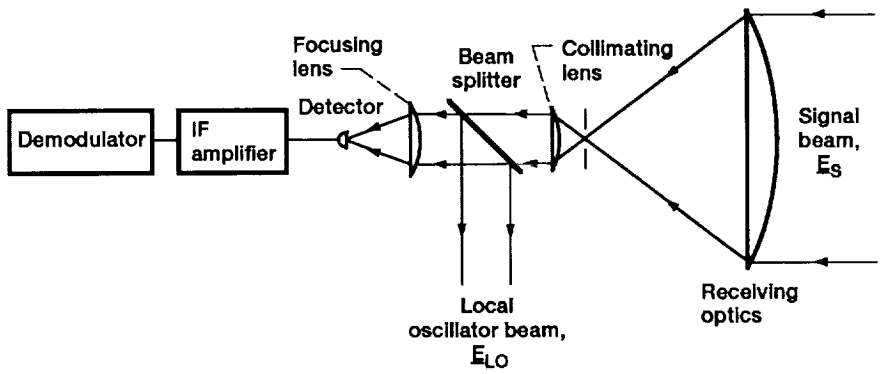


Figure 2.1.—Block diagram of a typical heterodyne receiver.

optical axis so that their focused images overlap on the detector surface. Misalignment of the two fields reduces the degree of photomixing, and thus, degrades the SNR, as shown in section 2.2.

2.1 Signal-to-Noise Ratio for the Heterodyne-Detection Receiver

Before discussing heterodyne optical detection, let us review optical-field coherence, since such coherence is necessary for heterodyning and homodyning to occur. Coherence is the ability of an optical field to interfere with itself or another field; there are two kinds of coherence: time, or temporal, coherence and spatial coherence.

Temporal coherence refers to a wave's phase at the same point in space, at different times. If the phase of a wave remains constant with time at some arbitrary point in space, the wave is fully time coherent. The coherence time is defined as the mean time between phase changes and is related to the bandwidth of the radiation $\Delta\nu$ (in Hz) by $\tau_{coh} = 1/(2\Delta\nu)$ sec. Coherence time also can be expressed in terms of the nominal wavelength λ (in meters) and linewidth $\Delta\lambda$ (in meters) as $\tau_{coh} = \lambda^2/(2c\Delta\lambda)$ sec, where c is the speed of light (in meters per second). The coherence length also can be defined as $c\tau_{coh}$ meters. Therefore, the more nearly monochromatic the radiation (i.e., the narrower the line width), the greater its temporal coherence. As an example, if we assume that an ordinary light bulb emits at a nominal wavelength of 0.7 μm over a wavelength range of 0.4 to 1.0 μm, then its line width is 0.6 μm and its bandwidth is about 367 000 GHz ($\Delta\nu \approx c\Delta\lambda/\lambda^2$); the corresponding coherence time and coherence length are 1.3 fs (1.3×10^{-15}) and 400 nm (400×10^{-9}), respectively. On the other hand, a helium-neon laser stabilized to operate on a single longitudinal mode in the laser cavity has a nominal wavelength of 0.6328 μm, a line width of 2×10^{-9} μm, and a bandwidth of about 1.5 MHz; in this case, the coherence time is 300 nsec and the coherence length is 100 m. Since long coherence time is important in heterodyne and homodyne detection, it is desirable to achieve a narrow emission line width by designing the laser to operate in a single longitudinal mode. (The effect of finite laser line width on system performance is considered in section 4.)

Spatial coherence is a measure of the optical field's phase coherence at two arbitrary points in space at the same time. The optical field is fully spatially coherent if the phase difference between any two points on a plane normal to the direction of propagation remains constant with time. Spatial coherence does not require temporal coherence; the phase at different points in space may fluctuate randomly in time, but so long as the phase difference between any two points remains constant, the field remains spatially coherent (i.e., the same fluctuation occurs at all points).

An imaginary test for spatial coherence (Young's two-slit interference experiment) would involve intercepting the optical field with an opaque screen having pin holes at the two arbitrary points and then studying the intensity distribution pattern on an observation screen. If the phase difference of the light emerging from the two pin holes remains constant, a stable pattern of light and dark interference fringes appears on the screen. If the field is not fully spatially coherent, the interference pattern may shift erratically or not be present at all. In the case of ordinary radiation (nonlaser) sources, the Van Cittert-Zernike theorem (ref. 5) provides the diameter of a circular area over which the radiation has a high degree of spatial coherence. This diameter is given by $d_{coh} = \lambda R/(2\pi\rho_b)$, where λ, R, and ρ_b are the respective mean wavelength, distance to the source, and radius of the source, which is assumed to be circular. For example, taking the central wavelength of the Sun to be 0.55 μm, we find that the light from the Sun is spatially coherent over only diameters of 0.02 mm on the Earth's surface. The light from a typical star, however, may have a coherent diameter of tens of centimeters (although its intensity is very small).

A laser can achieve perfect spatial coherence by oscillating on a single transverse mode within the laser cavity (a laser has both transverse and longitudinal modes of oscillation). Laser oscillation in the fundamental transverse mode—or the TEM_{00}, or gaussian, mode—results in a gaussian-intensity profile of the emitted radiation. Another advantage of gaussian-mode operation of the laser and the resulting spatial coherence is that the beam experiences minimal spreading from diffraction and, therefore, minimal divergence; a disadvantage, however, is that restriction to single-mode operation often reduces the total output power. Because heterodyning requires that the optical and received-signal fields be temporally and spatially coherent, heterodyne detection is also called coherent detection.

Because the optical field over the surface of a lens and the lens' diffracted field in the focal plane are related by a two-dimensional Fourier transform (appendix B), heterodyning of the received-signal and LO optical fields can be described in terms of the fields' diffraction patterns in the focal (detector) plane, or equivalently, in terms of plane waves at the receiver aperture.

Let us consider the received-signal and LO optical fields to be linearly polarized harmonic plane waves of the form

$$\begin{aligned} \vec{E}_S(\vec{r},t) &= \vec{A}_S \cos(\vec{k}_S \cdot \vec{r} - \omega_S t + \theta_S) \\ \vec{E}_L(\vec{r},t) &= \vec{A}_L \cos(\vec{k}_L \cdot \vec{r} - \omega_L t + \theta_L) \end{aligned} \tag{2.1}$$

where

$\vec{E}_S$, $\vec{E}_L$	signal and LO electric fields
$\vec{A}_S$, $\vec{A}_L$	field amplitudes
$\vec{k}_S$, $\vec{k}_L$	field propagation vectors ($\vec{k} = k_x\hat{i} + k_y\hat{j} + k_z\hat{k}$)
ω_S, ω_L	optical frequencies, rad/sec
θ_S, θ_L	signal and LO initial phases
$\vec{r}$	arbitrary position vector, $x\hat{i} + y\hat{j} + z\hat{k}$

Although not explicitly indicated, the amplitudes and phases of the received signal and LO fields may vary both with time and position. Because optical frequencies are so high, we may reasonably assume that amplitude or phase time variations occur over many optical periods. We may further satisfy ourselves that the expressions in equation (2.1) are those of plane waves if we recall the definition of a wave front: the surface over which a wave's phase, or the argument of the cosine function, is constant at an arbitrary point in time (say, at $t = 0$). Setting $t = 0$ and the cosine argument of either of the equations in equation (2.1) to a constant leads to

$$\vec{k} \cdot \vec{r} = xk_x + yk_y + zk_z = \text{constant} \tag{2.2}$$

which is simply the equation of a plane perpendicular to the propagation vector $\vec{k}$. Hence, the wave fronts are planar surfaces. As we move from plane to plane (at arbitrary points in time), $\vec{E}(\vec{r})$ varies sinusoidally with vectors $\vec{A}$ and $\vec{k}$, defining the electric field's plane of polarization. Because the wave has a spatial period equal to its wavelength, it can be shown that $|\vec{k}| = 2\pi/\lambda$. Finally, the term $-\omega t$ converts $|\vec{k}| = 2\pi/\lambda$ into a progressive plane wave moving in the positive k direction.

Since the equations in (2.1) represent electromagnetic waves, each satisfies the homogeneous (free-space) differential wave equation, which in right-handed Cartesian coordinates is given by

$$\nabla^2\psi = \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2} \tag{2.3}$$

The wave function (or in our case the $\vec{E}$-field) is $\psi(x,y,z,t)$ representing the disturbance in space and time, and the speed of propagation of the disturbance is v. Since the wave equation is a second-order, linear differential equation, the principle of superposition holds. Therefore, at a point where two or more optical fields overlap, the resultant field equals the vector sum of the individual fields. When this principle is applied to the heterodyne receiver, the resulting electric field in the aperture plane is

$$\vec{E}(\vec{r},t) = \vec{E}_S(\vec{r},t) + \vec{E}_L(\vec{r},t) \tag{2.4}$$

The intensity of the combined optical field is

$$I = \frac{1}{Z_o}\overline{E^2} = \frac{1}{Z_o}\overline{\vec{E}\cdot\vec{E}} \tag{2.5}$$

where the horizontal bar again denotes time averaging and $E^2 = \vec{E} \cdot \vec{E}$. Hence, we have

$$\begin{aligned} E^2 = \vec{E}\cdot\vec{E} &= (\vec{E}_S + \vec{E}_L)\cdot(\vec{E}_S + \vec{E}_L) \\ &= \vec{E}_S\cdot\vec{E}_S + \vec{E}_L\cdot\vec{E}_L + 2\vec{E}_S\cdot\vec{E}_L \\ &= E_S^2 + E_L^2 + 2\vec{E}_S\cdot\vec{E}_L \end{aligned} \tag{2.6}$$

The intensity is then

$$I = \frac{1}{Z_o}\left(\overline{E_S^2} + \overline{E_L^2} + \overline{2\vec{E}_S\cdot\vec{E}_L}\right) \tag{2.7}$$

If we assume that the difference between the optical carrier and LO frequencies is small in comparison to the frequencies themselves (i.e., $|f_S - f_L| \ll f_S$ or f_L) and that the time averaging interval T is long in comparison to the optical period but short in comparison to the "beat" period $1/(f_S - f_L)$, then taking the time average results in

$$\begin{aligned} I = \frac{A_S^2}{2Z_o} + \frac{A_L^2}{2Z_o} + \frac{1}{Z_o}\vec{A}_S\cdot\vec{A}_L\cos\Big[&(\vec{k}_S\cdot\vec{r}) - (\vec{k}_L\cdot\vec{r}) \\ &+ (\omega_S - \omega_L)t + (\theta_S - \theta_L)\Big] \\ = I_S + I_L + I_{IF} \end{aligned} \tag{2.8}$$

The total intensity is not simply the sum of the individual intensities, but it is modified by the last term, which is called the interference, or beat, term. If the received-signal and LO fields are not spatially and temporally coherent (i.e., if their phase angles fluctuate randomly and rapidly, independent of one another), then the interference term averages zero, and the total intensity is simply the sum of the separate intensities. Since heterodyne and homodyne optical receivers require an interference term, the received signal and LO optical fields must be mutually spatially coherent (i.e., $\theta_S - \theta_L$ must be constant or nearly so). Note that the beat term is centered at an intermediate frequency (IF) of $(\omega_S - \omega_L)$ and that it is sensitive to the relative polarization and propagation directions of the received signal and LO optical fields. In ideal heterodyne detection, both fields have the same polarization direction and are perfectly spatially aligned (i.e., $\vec{A}_S \cdot \vec{A}_L = A_S A_L$ and $\vec{k}_S = \vec{k}_L$), and the intensity becomes

$$I = I_S + I_L + \frac{A_S A_L}{Z_o}\cos\left[(\omega_S - \omega_L)t + (\theta_S - \theta_L)\right] \tag{2.9}$$

The significance and value of heterodyne detection is now apparent from equation (2.9). Note that, if the LO laser has a stable amplitude, frequency, and phase (i.e., if A_L, ω_L, and θ_L are constant), then the IF intensity term has an amplitude directly proportional to the amplitude of the received signal A_S, an IF linearly related to the received-signal frequency ω_S, and a phase linearly related to the received-signal phase θ_S. Therefore, unlike the case in direct detection, the received signal's amplitude, frequency, and phase are preserved in the intensity of the IF signal, permitting each of them to be modulated to carry information. Furthermore, in heterodyne detection, the IF signal power can be increased directly by increasing the LO laser power via A_L. The phase term indicates the need for

strict control of the transmission laser and LO-laser phase noise since phase variations will add directly to the received-signal's phase and will thus corrupt any phase modulation that may be impressed on the received signal. By multiplying equation (2.9) by the effective receiving aperture area, the power incident on the detector surface becomes

$$P(t) = A_r \frac{A_S^2}{2Z_o} + A_r \frac{A_L^2}{2Z_o} + A_r \frac{A_S A_L}{Z_o} \cos(\omega_{IF} t + \theta_{IF})$$

$$= P_S + P_L + 2\sqrt{P_S P_L} \cos(\omega_{IF} t + \theta_{IF}) \qquad (2.10)$$

where $\omega_{IF} = \omega_S - \omega_L$ and $\theta_{IF} = \theta_S - \theta_L$. After IF filtering and the removal of the dc terms, the desired IF signal current is

$$i_{IF}(t) = 2\sqrt{P_L P_S}\, \alpha e \cos(\omega_{IF} t + \theta_{IF}) \qquad (2.11)$$

where, as before, $\alpha = \eta/(h\nu)$. The desired signal power at the IF filter's output is then

$$S = \overline{i_{IF}(t)^2} = 2P_L P_S \alpha^2 e^2 \qquad (2.12)$$

The noise at the filter output consists of detector shot noise and circuit thermal noise. Shot noise power—which has components that result from the received-signal power P_S, LO power P_L, IF power P_{IF}, background power P_B, and detector dark current I_D, and which falls in the IF bandwidth B_{IF}—is given by the mean-square current, following equation (1.11),

$$\overline{i_s^2} = 2e^2 \alpha B_{IF}(P_S + P_L + P_B + P_{IF}) + 2eI_D B_{IF} \qquad (2.13)$$

The thermal-noise power at the filter output is, following equation (1.12),

$$\frac{4kTB_{IF}}{R_L} \qquad (2.14)$$

From equations (2.12) to (2.14), one obtains an output SNR of

$$SNR_{IF} = \frac{2P_L P_S \alpha^2 e^2}{2e^2 \alpha B_{IF}(P_S + P_L + P_B + P_{IF}) + 2eI_D B_{IF} + \dfrac{4kTB_{IF}}{R_L}} \qquad (2.15)$$

In the presence of increased LO power, LO shot noise, given by the term $2e^2 \alpha P_L B_{IF}$, can be made to dominate all other noise contributions in the denominator of equation (2.15). Thus, under a strong LO condition, the SNR becomes

$$SNR_{IF} = \frac{2P_L P_S \alpha^2 e^2}{2e^2 \alpha P_L B_{IF}} = \frac{\alpha P_S}{B_{IF}} = \frac{\eta P_S}{h\nu B_{IF}} \qquad (2.16)$$

which is the quantum-limited SNR bound for heterodyne detection at the output of the photodetector IF filter. This SNR is limited only by the shot noise produced by the LO. A second IF detection is required to extract the information signal from the IF subcarrier. To determine the demodulated SNR, we must consider the type of modulation used in impressing the baseband information signal onto the optical carrier. Rewriting the IF signal current in equation (2.11) in terms of the electric-field amplitudes gives

$$i_{IF}(t) = A_S A_L K \cos(\omega_{IF} t + \theta_{IF}) \qquad (2.17)$$

where K is a constant given by $K = A_r \alpha e / Z_o$. If amplitude modulation (AM) is used to modulate the information signal $m(t)$ onto the optical carrier, then the IF signal current takes the form

$$i_{IF}(t) = \frac{A_S}{2}[1 + m(t)] A_L K \cos(\omega_{IF} t) \qquad (2.18)$$

where spatial coherence is assumed between the received-signal and LO optical fields such that $\theta_{IF} = 0$ and that $|m(t)| \leq 1$ to prevent overmodulation. Recall that for coherent product demodulation of an AM, or of a double sideband-suppressed carrier (DSB/SC) signal, an improvement factor of 2 occurs in going from the IF SNR to the baseband SNR (ref. 6). This improvement factor results from the coherent addition of the signal components in the sidebands while the noise components add noncoherently. This improvement factor also occurs in envelope detection if the IF SNR is large. Therefore, if we assume that the information signal has maximum power $P_m = 1$ and bandwidth B_m and that the IF bandwidth is set to $B_{IF} = 2B_m$, the demodulated SNR for a strong laser local oscillator is

$$SNR_D = 2\, SNR_{IF} = 2\left(\frac{\eta P_S}{h\nu 2B_m}\right) = \frac{\eta P_S}{h\nu B_m} \qquad (2.19)$$

When we compare this SNR with the quantum-limited SNR for baseband direct detection in equation (1.15), we see that AM with ideal heterodyne detection yields a 3-dB improvement in SNR. It might also be noted that AM with homodyne detection yields the same SNR as does equation (2.19) while eliminating the need for AM demodulation following photodetection, since the (low-pass) filtered detector output contains $m(t)$ at baseband.

For FM, the IF signal will have the form

$$i_{IF}(t) = A_S A_L K \cos[\omega_{IF} t + \theta_{IF}(t)] \qquad (2.20)$$

where

$$\theta_{IF}(t) = 2\pi\Delta f \int_{t_o}^{t} m(\tau)\, d\tau$$

and the peak frequency deviation from the IF frequency is Δf. By Carson's rule, the modulated IF carrier occupies a bandwidth of

$$B_{IF} = 2(\beta + 1)B_m \tag{2.21}$$

where $\beta = \Delta f / B_m$ is the FM index. FM demodulation of the IF signal can be performed by a limiter-discriminator detector that yields a demodulated SNR for strong laser LO, given by

$$SNR_D = 3\beta^2 \frac{B_{IF}}{2B_m} SNR_{IF} = \frac{3}{2}\beta^2 \frac{\eta P_S}{h\nu B_m} \tag{2.22}$$

To achieve this FM improvement, the IF SNR at the input of the FM demodulator must exceed the FM threshold value of 12 dB. Thus, the output SNR can be increased by using large values of the modulation index (i.e., wideband FM operation), which results in a penalty of increased required bandwidth.

2.2 Signal-to-Noise Ratio Degradation Resulting From Received Signal Field and Local Oscillator Field Misalignment

The heterodyne optical-field intensity described by equation (2.9) assumes that the received-signal and LO optical field have identical polarization vectors, identical propagation vectors normal to the detector surface (i.e., the fields are perfectly spatially aligned), and identical, exactly overlapping diffraction patterns on the detector surface. The photo mixing process takes place in the region of overlap on the detector surface. However, ideal heterodyning conditions, especially perfect alignment and matching diffraction patterns, are extremely difficult to achieve in real systems. Consequently, the resultant SNR is degraded from the ideal value obtained in equation (2.16).

If one assumes that both the received-signal field and the LO field can be approximated as plane waves, each will have a diffracted optical field in the detector plane of the form (appendix B)

$$E(r) = \left(-je^{jk_p f} e^{\frac{jk_p r^2}{2f}}\right) \frac{\pi a^2}{\lambda f} A_o \frac{2J_1(k_p ar/f)}{(k_p ar/f)} \tag{2.23}$$

where

k_p propagation number, $2\pi/\lambda$

f focal length of the receiving telescope

r radial distance from the pattern center

a radius of the receiving aperture

A_o amplitude of the incident field on the aperture

The magnitude of this field is

$$|E(r)| = \frac{\pi a^2}{\lambda f} A_o \left| \frac{2J_1(k_p ar/f)}{(k_p ar/f)} \right| \tag{2.24}$$

Since intensity is proportional to the squared magnitude of the electric field, the intensity distribution in the focal plane is

$$I(r) = \frac{|E(r)|^2}{2Z_o} = \frac{A_o^2}{2Z_o} \left(\frac{\pi a^2}{\lambda f}\right)^2 \left[\frac{2J_1(k_p ar/f)}{(k_p ar/f)}\right]^2 \tag{2.25}$$

Since $I_o = A_o^2/(2Z_o)$ is the intensity of the incident plane wave, and the maximum on-axis ($r = 0$) intensity in the focal plane is $I_{\max} = I_o\,(\pi a^2/\lambda f)^2$, equation (2.25) may be written as

$$I(r) = I_{\text{MAX}} \left[\frac{2J_1(2\pi ar/\lambda f)}{(2\pi ar/\lambda f)}\right]^2 \tag{2.26}$$

where we have made the substitution $k = 2\pi/\lambda$. Note that the total power in the diffracted field is

$$\begin{aligned}
\int_A I(r)dA &= \int_{\theta=0}^{\theta=2\pi} \int_0^{r=\infty} I(r) r\, dr\, d\theta \\
&= 2\pi I_{\text{MAX}} \int_0^\infty \frac{[2J_1(k_p ar/f)]^2}{(k_p ar/f)^2} r\, dr \\
&= \frac{8\pi I_o (\pi a^2/\lambda f)^2}{(2\pi a/\lambda f)^2} \int_0^\infty \frac{J_1^2(k_p ar/f)}{r} dr \\
&= 2\pi a^2 I_o \left(\frac{1}{2}\right) = A_r I_o = P_o
\end{aligned} \tag{2.27}$$

where A_r and P_o are the receiver aperture area and the incident power, respectively. Therefore, the power in the focused field equals the received power over the aperture area—the result that one would expect from the conservation of energy.

The electric-field distribution in equation (2.27) is for a plane wave that arrives normal to the receiver aperture so that its diffraction pattern is centered on the detector surface. If we assume that the focal plane is described by an x-y coordinate plane with the detector surface centered on the origin and that the detector is a rectangle of width w along the x-axis and height h_d along the y-axis, then a plane wave normally arriving will have the center of its focused pattern at the origin (i.e., the center of the detector surface). A plane wave not arriving normally will have a focused field pattern of the same form as equation (2.26), but with its center displaced from the origin. Specifically, if the off-axis arrival angles measured from the normal are θ_x and θ_y, then the focused field is centered at

$$\left.\begin{aligned} x_c &= f - \tan(\theta_x) \\ y_c &= \sqrt{f^2 + x_c^2}\tan(\theta_y) \approx f - \tan(\theta_y) \end{aligned}\right\} \tag{2.28}$$

if we assume that the detector dimensions are much smaller than the focal length f. In the focal plane (x,y) coordinates, the E-field magnitude has the form

$$|E(x,y)| = \left| \frac{2J_1(k_p a r_o/f)}{(k_p a r_o/f)} \right| \quad r_o = \sqrt{(x - x_c)^2 + (y - y_c)^2} \tag{2.29}$$

Since the focus here is on the field pattern variation, the constant multiplying factor is ignored for the moment.

Reconsidering the optical heterodyne receiver, let the received field have wavelength λ_R and arrive off-axis from the normal by angles θ_{xR} and θ_{yR}. Similarly, let the local field have wavelength λ_L and arrive off-axis from the normal by angles θ_{xL} and θ_{yL}. If we ignore the constant factor, the magnitudes of the two fields in the detector plane are then

$$\left.\begin{aligned} |E_R(x,y)| &= \left| \frac{2J_1(k_R a r_R/f)}{k_R a r_R/f} \right| \\ |E_L(x,y)| &= \left| \frac{2J_1(k_L a r_L/f)}{k_L a r_L/f} \right| \end{aligned}\right\} \tag{2.30}$$

where

$$k_R = 2\pi/\lambda_R \quad k_L = 2\pi/\lambda_L \quad r_R = \sqrt{(x - x_R)^2 + (y - y_R)^2}$$

$$x_R = f\tan(\theta_{xR}) \quad y_R = f - \tan(\theta_{yR})$$

$$r_L = \sqrt{(x - x_L)^2 + (y - y_L)^2} \quad x_L = f\tan(\theta_{xL})$$

$$y_L = f - \tan(\theta_{yL})$$

Since the LO optical field pattern is not precisely superimposed on the received-signal field pattern, not all the LO power will contribute to useful IF signal power. As with time waveforms, we can define a spatial correlation coefficient or a field alignment efficiency that measures how closely the two field patterns match or overlap on the detector surface as

$$\eta_{fa} = \frac{\int_{-h/2}^{h/2}\int_{-w/2}^{w/2} |E_R(x,y)|\,|E_L(x,y)|\,dx\,dy}{\int_{-h/2}^{h/2}\int_{-w/2}^{w/2} |E_{Rc}(x,y)|\,|E_{Lc}(x,y)|\,dx\,dy} \tag{2.31}$$

where E_{Rc} and E_{Lc} represent field patterns exactly centered on each other. For a specified set of misalignment angles, this ratio of integrals yields a value between zero and 1 and acts as an SNR-suppression factor. Note that from equations (2.30), as the receiving aperture size increases, the focused field patterns decrease and the SNR degradation for a given field misalignment increases (i.e., η_{fa} decreases).

An approximation of equation (2.31) can be arrived at by assuming $\lambda_L \approx \lambda_R$ and considering only each pattern's circular main lobe (i.e., its Airy disc, or focal spot). This disc has a radius of $1.22\lambda f/D$, where $D = 2a$ is the receiver aperture diameter and f is the focal length (appendix B). The substitution of $r = 1.22\,\lambda f/D$ for the upper limit of r in the surface integral in equation (2.27) reveals that about 83.8 percent of the total power is contained within the Airy disc. For an approximation of equation (2.31), the overlap area of the two Airy discs is computed on the detector surface. The ratio of this overlap area to the LO disc area is approximately the fraction of total LO power that contributes to heterodyning. The overlap area of two overlapping circles of equal diameter and radius R, whose centers are separated by a distance of d, is given by

$$A = 2\left(R^2 \cos^{-1}\frac{d}{2R} - \frac{d}{2}\sqrt{R^2 - \frac{d^2}{4}}\right) \tag{2.32}$$

The alignment efficiency is then approximately

$$\eta_{Fa} = \frac{A}{\pi R^2} \quad R = 1.22\lambda\frac{f}{d} \quad d = f\tan\theta \tag{2.33}$$

where θ is the misalignment angle between the two fields. If the angular radius of each disc is defined as $\theta_o = 1.22\lambda/D$ rad, then equation (2.33) can be expressed as

$$\eta_{Fa} = \frac{2}{\pi}\cos^{-1}\left[\frac{\tan\theta}{2\theta_o}\right] - \frac{\tan\theta}{\pi\theta_o^2}\sqrt{\theta_o^2 - \tan^2\left(\frac{\theta}{4}\right)} \tag{2.34}$$

To avoid significant misalignment losses, the angle between the two fields generally must be no greater than the angular radius θ_o of the Airy disc of either diffraction pattern (which also is the threshold for the Rayleigh criterion for resolving the images of two point sources in the focal plane of a lens). A misalignment angle of this value would correspond to the central maximum or center, of one intensity pattern occurring at the first null, or disc edge, of the other. In this case, $\theta = \theta_o$ and equation (2.34) yield a loss of about 0.40, or -4 dB. Table 2.1 depicts the misalignment loss, or alignment efficiency, as a function of the misalignment angle, normalized to the diffraction-limited angle, as computed from equation (2.34).

By accounting for the difference in the polarization directions of the signal and LO fields, the overall heterodyning efficiency, or SNR degradation, may be defined as

$$\eta_{\text{het}} = \eta_{fa}\eta_{\text{pol}} = \eta_{fa}\cos^2\varphi \tag{2.35}$$

where φ is the angle between the polarization vectors (i.e., the angle between $\vec{A}_S$ and $\vec{A}_L$). Normally, misalignment error between the two fields is the dominant contributor to overall SNR degradation.

TABLE 2.1.—BEAM MISALIGNMENT LOSS AS A FUNCTION OF NORMALIZED MISALIGNMENT ANGLE

θ/θ_0 (where $\theta_0 = 1.22\ \lambda/d$)	Loss, dB
0.0	0.0
.1	−.286
.2	−.590
.4	−1.266
.6	−2.049
.8	−2.970
1.0	−4.078
1.2	−5.455
1.4	−7.256
1.6	−9.826
1.8	−14.273

3.0 Derivation of Bit-Error-Rate for Digital Direct-Detection M-ary Pulse Position Modulation

Here, the bit error rate (BER) of a direct-detection receiver for an optical system transmitting digital information is determined. The system is assumed to use an APD detector and M-ary pulse position modulation (PPM). Although other pulse-modulation techniques, such as pulse polarization modulation or on-off keying (OOK), are possible with direct detection, PPM is the most commonly used format because multiple bits of information can be transmitted per optical pulse. The objective here is to determine the required signal power needed at the DD receiver in order to achieve a specified BER, given certain system parameters (e.g., data rate, PPM order, and background interference), and the physical parameters of the receiver (e.g., APD gain, quantum efficiency, and excess noise factor).

In an M-ary PPM digital system, the transmitting laser is pulsed on and off at a prescribed pulse repetition frequency (PRF), producing an optical pulse of width τ_p sec and peak power P_p, every $T = 1/\text{PRF}$ sec. A PPM modulator driving the laser accepts k_b bits at a time from the input data stream and sends the information in each k-bit symbol by delaying the optical pulse into one of $M = 2^{k_b}$ time slots. The width of each time slot is $\tau \approx \tau_p$ seconds. Since each PPM frame or word carries $k_b = \log_2 M$ bits of information and has a duration of $T = M\tau$ seconds, the channel data rate is

$$R_b = \frac{k_b}{T} = \frac{\log_2 M}{M\tau} \quad \text{bps} \tag{3.1}$$

The average laser power P_a is related to the peak power by

$$P_p\tau = P_aT = \frac{P_a}{PRF} \tag{3.2}$$

Figure 3.1 provides an example of a pulse position modulation (PPM) encoded bit stream for $M = 2$, binary PPM (BPPM), and $M = 4$, quaternary PPM (QPPM). After the receiver has established PPM frame and slot synchronization (typically, using Manchester PCM encoding as an aid in clock recovery), it must decide which of the M time slots contains the optical signal pulse. In this case, the optimum-detection strategy is maximum-count detection, because it is a maximum-likelihood scheme. In maximum-count detection, the detector output current is integrated over each of the M time slots, and the slot with the maximum photoelectron count is selected as the one in which the optical pulse was transmitted. The integration may be carried out sequentially over each time slot, using a single integrator, or in parallel, using a bank of integrators with appropriate delays and pairwise comparisons of the slot counts. After the PPM demodulator has selected the slot with the maximum-count, it outputs k data bits of the corresponding M-ary symbol. This maximum-count scheme outperforms threshold detection, wherein the photoelectron count in each slot must be compared against a threshold value, which further must be varied in accordance with background-noise fluctuations.

Referring to equation (1.16), which gave the direct-detection SNR with photomultiplication at the output of a baseband filter of bandwidth B_m, we will determine the SNR occurring within a PPM time slot. The mean and variance of the photoelectron-count distribution for both signal and nonsignal time slots will be determined by approximating the detector output current as a gaussian random variable. (A signal slot is a slot that contains the optical pulse as well as noise, whereas a nonsignal slot contains only noise.)

If we let P_s in the numerator of equation (1.16) represent the received pulse power, the desired-signal power at the detector-preamplifier output during a signal slot interval is

$$S = (\bar{G})^2\alpha^2e^2P_s^2 = (\bar{G})^2I_s^2 = (\bar{G}I_s)^2 \tag{3.3}$$

where $\bar{G}I_s$ is the photomultiplied signal current. (Note that, since this is now a digital system, the P_m factor is ignored.)

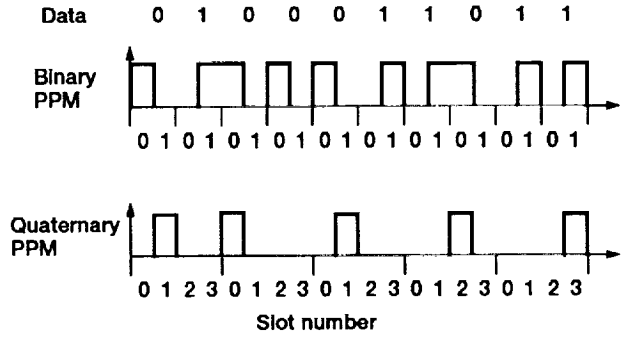


Figure 3.1.—Sample bit stream in pulse position modulation (PPM) format.

If we factor out $2B_m$ from the denominator of equation (1.16), the two-sided power spectral density of the noise in the detector-preamplifier output current is

$$G_x(f) = \overline{G^2}e^2\alpha P_s + \overline{G^2}e^2\alpha P_B + \overline{G^2}eI_D + \frac{2kT_{eq}}{R_L}$$

$$= \overline{G^2}eI_s + \overline{G^2}eI_B + \overline{G^2}eI_D + \frac{2kT_{eq}}{R_L} \qquad (3.4)$$

Again, these terms represent photomultiplied-signal shot noise, background shot noise, dark-current shot noise, and preamplifier thermal noise, respectively. For an APD, the dark current, which is the current generated when no light is incident on the device, is in truth a combination of a bulk dark current and a surface dark current. The bulk dark current I_{DB} arises from electrons or holes that are thermally generated in the p-n junction of the device, therefore undergoing avalanche gain. The surface dark current I_{DS} is a leakage current unaffected by avalanche gain. It is a function of surface defects, bias voltage, and surface area, among other things (ref. 7). Although the shot noise produced by either of these currents usually is negligible in comparison with signal- or background-generated shot noise or circuit thermal noise, it is considered here for reasons of completeness. For a typical silicon APD, the bulk and surface dark currents might be 0.1 and 10.0 nA, respectively. Upon breaking the dark current into a gain-dependent bulk component and a gain-independent surface component, equation (3.4) can be rewritten as

$$G_x(f) = \overline{G^2}(eI_S + eI_B + eI_{DB}) + eI_{DS} + \frac{2kT_{eq}}{R_L} \quad \text{W/Hz} \qquad (3.5)$$

Figure 3.2 illustrates the APD-equivalent circuit model and typical parameter values. Equations (3.3) and (3.5) refer to the signal and noise power at the input of the PPM time-slot integrator. Since the integrator integrates current over an integration time of τ seconds, the output of the integrator represents the total number of photoelectrons that result from both signal and noise within a slot interval. The count distribution is completely characterized by the mean and variance as a result of treating the photoelectron count as a gaussian, ergodic, random process. During signal and nonsignal slot times, the mean value of the integrator output is, respectively,

$$\left.\begin{aligned} \mu_s &= \int_0^\tau \bar{G}I_s\, dt = \bar{G}I_s\tau \\ \mu_n &= 0 \end{aligned}\right\} \qquad (3.6)$$

The variance equals the average power in the noise at the integrator output, which follows from

$$\sigma_x^2 = E[x^2] - \mu_x^2 \qquad (3.7)$$

where $E[\ ]$ denotes the expected value operator. Here, for a random signal $x(t)$, $E[X^2]$ is the mean square value equal to the total average power of the signal (dc power + ac power), whereas μ_x^2 is the power in the dc component alone. Therefore, since the input noise power spectral density is $G_x(f)$, the total output noise power, or variance, for a signal slot is given by

$$\sigma_s^2 = \int_{-\infty}^{+\infty} G_y(f)\, df \qquad (3.8)$$

Excess noise factor, F, dB	5
Quantum efficiency, η	0.85
Gain-independent dark current, I_{DS}, nA	10
Gain-dependent dark current, I_{DB}, nA	0.1
Intrinsic capacitance, $C_A + C_Q$, pF	0.1
Thermal noise temperature, T_{EQ}, K	290
APD gain, G	200
Receiver load resistor, R_L	bandwidth driven

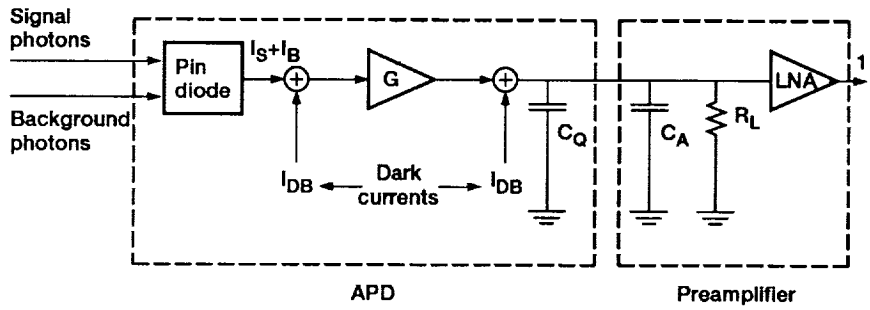


Figure 3.2.—Avalanche photodiode (APD) circuit model with high-impedance preamplifier.

where $G_y(f)$, the output-noise power spectral density, is equal to

$$G_y(f) = |H(f)|^2 G_x(f) \tag{3.9}$$

The Fourier transform of the impulse response of the integrator $h(t)$ is $H(f)$. Since a unit impulse has unit area and since the integration time is equal to the slot time τ, the impulse response of the integrator is simply

$$\begin{aligned} h(t) &= 1 \quad 0 \le t \le \tau \\ &= 0 \quad \text{all other cases} \end{aligned} \tag{3.10}$$

Taking the Fourier transform of $h(t)$, we arrive at

$$H(f) = \tau \frac{\sin(\pi f \tau)}{\pi f \tau} e^{-j\pi f \tau} \tag{3.11}$$

$$|H(f)|^2 = \tau^2 \frac{\sin^2(\pi f \tau)}{(\pi f \tau)^2} \tag{3.12}$$

Since $G_x(f)$ is a flat-power spectral density constant over all frequencies, we obtain

$$\begin{aligned} \sigma_s^2 &= G_x(f) \int_{-\infty}^{\infty} |H(f)|^2 \, df \\ &= \tau G_x(f) \\ &= \tau e \left(\overline{G^2} I_s + \overline{G^2} I_B + \overline{G^2} I_{DB} + I_{DS} + \frac{2kT_{eq}}{eR_L} \right) \end{aligned} \tag{3.13}$$

The variance in the photoelectron count during a nonsignal slot σ_n^2 is given by equation (3.13) with $I_s = 0$

$$\sigma_n^2 = \tau e \left(\overline{G^2} I_B + \overline{G^2} I_{DB} + I_{DS} + \frac{2kT_{eq}}{eR_L} \right) \tag{3.14}$$

Let us define the average signal photoelectron count per signal time slot interval as

$$K_S = \frac{I_s}{e}\tau = \alpha P_S \tau = \frac{\eta}{h\nu} P_S \tau \tag{3.15}$$

The mean of the integrator output during a signal slot can then be expressed as $m_s = \bar{G} I_s \tau = \bar{G} K_s e$. The SNR (i.e., average signal power divided by the total average noise power) at the integrator output, within a signal slot interval, is then

$$\begin{aligned} \text{SNR} &= \frac{m_s^2}{\sigma_s^2} \\ &= \frac{\overline{G^2} K_S^2 e^2}{\overline{G^2} \tau e I_s + \overline{G^2} \tau e I_B + \overline{G^2} \tau e I_{DB} + \tau e I_{DS} + \dfrac{\tau 2kT_{eq}}{R_L}} \\ &= \frac{K_S^2}{\dfrac{\overline{G^2}}{\overline{G}^2} I_S \dfrac{\tau}{e} + \dfrac{\overline{G^2}}{\overline{G}^2} I_B \dfrac{\tau}{e} + \dfrac{\overline{G^2}}{\overline{G}^2} I_{DB} \dfrac{\tau}{e} + \dfrac{1}{\overline{G}^2} I_{DS} \dfrac{\tau}{e} + \dfrac{1}{\overline{G}^2} \dfrac{\tau}{e^2} \dfrac{2kT_{eq}}{R_L}} \\ &= \frac{K_S^2}{FK_S + FK_B + FK_{DB} + K_{DS} + K_T} \end{aligned} \tag{3.16}$$

where F is the APD excess noise factor, and the noise photoelectron counts caused by background radiation, bulk dark current, surface dark current, and thermal noise, respectively, are given by

$$K_B = \frac{I_B}{e}\tau \quad K_{DB} = \frac{I_{DB}}{e}\tau \quad K_{DS} = \frac{I_{DS}}{e}\frac{\tau}{\bar{G}^2} \quad K_T = \frac{2kT_{eq}\tau}{e^2 \bar{G}^2 R_L} \tag{3.17}$$

The denominator of equation (3.16) can be expressed as the sum of a signal-dependent noise component FK_S and a signal-independent noise component K_N given by

$$K_N = F(K_B + K_{DB}) + K_{DS} + K_T \tag{3.18}$$

The SNR can therefore be written as

$$SNR = \frac{K_S^2}{FK_S + K_N} \tag{3.19}$$

Note that in the absence of all background, dark current, and thermal noise, the $SNR = K_S/F$ and is bounded by signal-generated shot noise and APD gain noise. It also should be noted that the excess noise factor F for an APD is related to the average gain $\bar{G}$ by

$$F = k_{\text{eff}} \bar{G} + (1 - k_{\text{eff}}) \left(2 - \frac{1}{\bar{G}} \right) \tag{3.20}$$

where k_{eff} is the ratio between the hole and electron-ionization coefficients, or the effective ionization ratio. The smaller the ionization ratio, the better, since avalanche noise is lowest in APD's where the gain process is generated primarily by the

highest-mobility carriers. A typical value of k_{eff} is 0.01, although values as low as 0.007 have been achieved.

As expressed in equation (3.17), the thermal noise count K_T is time-slot dependent. From equation (3.17), it appears that K_T can be made negligible simply by making the preamplifier load resistance R_L large enough. However, to allow good pulse reproduction and to prevent the smearing of pulse energy into adjacent PPM time slots, the detector-preamplifier circuit bandwidth W_f must be set approximately equal to the pulse bandwidth, $W_P \approx 1/\tau$. If we let $C = C_Q + C_A$ be the total intrinsic capacitance of the detector-preamplifier circuit (fig. 3.2), then the circuit can be approximated by a simple low-pass resistor-capacitor (RC) filter whose 3-dB bandwidth is $W_f = 1/(2\pi R_L C)$. Therefore, R_L affects both the receiver bandwidth and the thermal-noise count K_T. Although increasing R_L reduces the thermal-noise count, it also reduces the receiver bandwidth. For reasonably good pulse fidelity, we want

$$W_f = \frac{1}{2\pi R_L C} \approx W_p = \frac{1}{\tau} \Rightarrow \tau \approx 2\pi R_L C \tag{3.21}$$

If we now assume that, as the time slot width τ changes, R_L is adjusted so as to always satisfy equation (3.21), then the thermal-noise count can be expressed as

$$K_T = \frac{2kT_{eq}\tau}{e^2\bar{G}^2R_L} = \frac{2kT_{eq}(2\pi R_L C)}{e^2\bar{G}^2R_L} = \frac{4\pi kT_{eq}C}{e^2\bar{G}^2} \tag{3.22}$$

which is time-slot independent. Therefore, the photoelectron counts resulting from signal, background radiation, and dark current are time-slot dependent, whereas under the assumption in equation (3.21), the thermal noise count can be considered to be time-slot independent (ref. 8). This distinction is important because, as will be shown later, the BER performance of M-ary PPM depends on whether the dominant noise contributor is time-slot dependent (i.e., τ-dependent) or time-slot independent (i.e.,τ -independent).

In terms of signal and noise photoelectron counts, the means and variances of the count distribution (gaussian) for signal and nonsignal PPM time slots at the integrator output are

$$\begin{aligned} m_s &= \bar{G}K_S e \qquad & \sigma_s^2 &= e^2(\bar{G})^2(FK_S + K_N) \\ m_n &= 0 \qquad & \sigma_n^2 &= e^2(\bar{G})^2K_N \end{aligned} \tag{3.23}$$

If we let the photoelectron count at the integrator output be denoted by the random variable X, then the corresponding gaussian probability densities are

$$\left.\begin{aligned} p_s(x) &= N(m_s,\sigma_s^2) = \frac{1}{\sqrt{2\pi}\sigma_s} e^{\left[-(x-m_s)^2/2\sigma_s^2\right]} \\ p_n(x) &= N(m_n,\sigma_n^2) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{\left(-x^2/2\sigma_n^2\right)} \end{aligned}\right\} \tag{3.24}$$

The integrator will produce a photoelectron count for each of the M time slots. One of these time slots will contain the signal pulse, whereas the remaining $M-1$ slots will contain only noise. If we let s be the photoelectron count of the signal slot, then the probability that the photoelectron count n in an arbitrary nonsignal slot is less than s is given by

$$\begin{aligned} Prob(n<s) &= \int_{-\infty}^{s} p_n(x)\,dx \\ &= \int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}\sigma_n} e^{\left(-x^2/2\sigma_n^2\right)}dx \\ &= 1 - Q(s/\sigma_n) \end{aligned} \tag{3.25}$$

where $Q(x)$ is the "communications" complementary error function, defined as the area under the standard normal density (i.e., zero mean and unit variance) from $u = x$ to infinity

$$Q(x) \equiv \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{(-u^2/2)}\,du \tag{3.26}$$

The probability that the signal-slot count is the maximum count over all slots is equal to the probability that the counts of all $M-1$ nonsignal slots are less than s

$$Prob\ (\text{all } n < s) = \left(1 - Q\left[\frac{s}{\sigma_n}\right]\right)^{M-1} \tag{3.27}$$

Equation (3.27) gives the probability of correctly deciding a PPM word for a specific value s of the signal-slot count. Note, however, that the signal-slot count is a gaussian random variable. Therefore, equation (3.27) must be averaged over all values of s so that the probability of a correct PPM word (PWC) decision is

$$\begin{aligned} PWC &= \int_{-\infty}^{\infty} \left(1 - Q\left[\frac{s}{\sigma_n}\right]\right)^{M-1} p_s(s)\,ds \\ &= \int_{-\infty}^{\infty} \left(1 - Q\left[\frac{s}{\sigma_n}\right]\right)^{M-1} \frac{1}{\sqrt{2\pi}\sigma_s} e^{\left[-(s-m_s)^2/2\sigma_s^2\right]}ds \end{aligned} \tag{3.28}$$

If we make a change in the variable and let $u = (s - m_s)/\sigma_s$, equation (3.28) becomes

$$PWC = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(1 - Q\left(\frac{m_s + u\sigma_s}{\sigma_n}\right)\right)^{M-1} e^{(-u^2/2)}du \tag{3.29}$$

The probability of a PPM word error (PWE) is then simply $1 - PWC$. Since PPM consists of M orthogonal signals, the bit error rate, BER = P_B, is related to the word (or symbol) error rate by

$$\frac{P_B}{PWE} = \frac{\text{number of ways bit error can occur}}{\text{number of ways word error can occur}} = \frac{M/2}{M-1} \tag{3.30}$$

This equation follows from the fact that, for a given transmitted PPM word, there are $M - 1$ incorrect words in the signal set, and for any bit position, half the words in the set will contain the wrong bit. As M increases, the BER will approach half the word error rate. From equations (3.29) and (3.30), the BER for M-ary PPM direct detection using maximum count detection is therefore

$$P_B = \frac{1}{2}\frac{M}{M-1}\left\{1 - \frac{1}{\sqrt{2\pi}} \times \int_{-\infty}^{\infty}\left[1 - Q\left(\frac{m_s + u\sigma_s}{\sigma_n}\right)\right]^{M-1} e^{(-u^2/2)}du\right\} \tag{3.31}$$

Since m_s, σ_s, and σ_n are related to the signal and noise photoelectron counts through equation (3.23), equation (3.31) can be used to determine the signal photoelectron count and, therefore, the power necessary to achieve a particular BER, given PPM order M, bit rate R, background power P_B, and the detector and preamplifier parameters. Unfortunately, equation (3.31) cannot be expressed in closed form; however, for $M = 2$ (BPPM), equation (3.31) can be simplified as

$$P_{B_{BPPM}} = Q\left[\sqrt{\frac{m_s^2}{\sigma_s^2 + \sigma_n^2}}\right] = Q[\sqrt{\gamma}] \tag{3.32}$$

where (eqs. (3.14) and (3.16))

$$\gamma = \frac{K_S^2}{FK_S + 2FK_B + 2FK_{DB} + 2K_{DS} + 2K_T} \tag{3.33}$$

Once K_S has been found for a desired BER, the corresponding, required received peak-pulse power can be found from equations (3.1) and (3.15) as

$$P_s = \frac{K_S}{\alpha\tau} = \frac{K_S}{\alpha}\frac{MR}{\log_2 M} = K_S\frac{h\nu}{\eta}\frac{MR}{\log_2 M} \tag{3.34}$$

The required, received average power is then simply

$$P_r = P_s\frac{\tau}{T} = P_s\frac{\tau}{M\tau} = \frac{P_s}{M} = K_S\frac{h\nu}{\eta}\frac{R}{\log_2 M} \tag{3.35}$$

From equation (3.31), for a given set of detector and preamplifier parameters (η, $\bar{G}$, F, T_{eq}, C, I_{DS}, I_{DB}), we find that the BER performance of M-ary PPM depends on K_S, K_N, and M. The signal-independent noise component K_N, in turn, depends on the τ-dependent noise counts (K_B, K_{DB}, K_{DS}) and on the τ-independent noise count K_T (R_L is assumed to be adjusted with slot width so as to always satisfy equation (3.21)). Finally, the slot width τ is related to the data rate R and the PPM order M through equation (3.1).

For a given data rate, the BER performance for different values of M depends on whether the dominant noise is time-slot-dependent background or dark-current noise or time-slot-independent thermal noise. For example, if background noise is dominant, then for a given signal count, BER improves as M increases (fig. 3.3) because fewer background noise photoelectrons occur in the narrower slot time as M increases. Likewise, for a given BER, the required signal count K_S decreases as M increases. From equation (3.35), we observe that the required average power also will be reduced at a higher PPM order (R is assumed constant). The calculation of equation (3.34), however, shows that the required peak power increases with increasing M (since the change in K_S is not as great as the change in τ).

On the other hand, if thermal noise dominates, then for a given signal count, the BER worsens as M increases (figure 3.4). Likewise, the required signal count to maintain a given BER increases as M increases. This is because the thermal noise count K_T remains nearly constant, regardless of the slot width, whereas the probability of making a PPM word error increases as M increases because there are more word slots. Figure 3.4 illustrates, however, that K_S changes slowly with M, so that as M increases, the required average power is

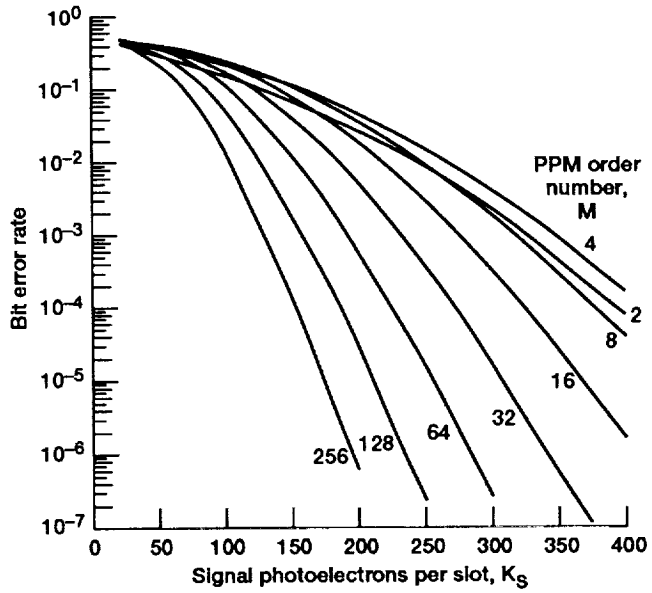


Figure 3.3.—Bit error rate (BER) versus signal photocounts per slot for M-ary PPM. (Dominant noise is background noise).

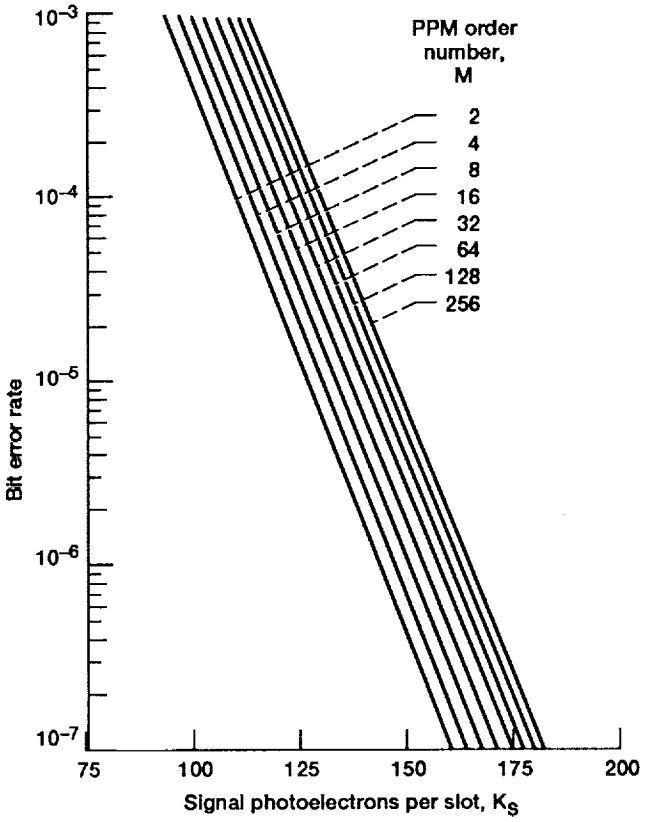


Figure 3.4.—Bit error rate (BER) versus signal photocounts per slot for M-ary PPM (dominant noise is thermal noise).

reduced while the required peak power is increased, for a constant data rate.

Consequently, using a high-order PPM reduces the average power requirement along with the transmitter laser pulse repetition frequency (PRF). Disadvantages of such use include higher peak-power requirements, larger receiver bandwidth, and greater modulator and demodulator complexity. The optimal choice for M in a particular application will depend on the type of transmitter-laser to be used and on receiver-bandwidth limitations. For example, semiconductor lasers are not suitable for high M-ary PPM systems because of their low peak-to-average power ratio, whereas, for high M-ary systems, Nd:YAG lasers often are used.

4.0 Bit-Error-Rate Performance of Heterodyne Optical Detection

This section considers the BER performance of heterodyne detection using various digital modulation methods. Heterodyne or homodyne systems can offer a significantly greater receiver sensitivity than can direct-detection systems, which means, in effect, that a given BER can be achieved with less received signal power (or with fewer signal counts per bit). Such systems also offer much higher background-noise rejection, shown in section 5. Greater receiver sensitivity is contingent on two factors: (1) the received-signal and LO optical fields must be polarization matched and spatially aligned so that their focused patterns overlap on the detector surface; (2) the two fields must have a high degree of spatial and temporal coherence. The quantum-limited BER can only be achieved under these conditions (section 4.1), and phase noise in either laser can seriously degrade BER performance (section 4.2). Consequently, optical sources must have high-frequency stability and narrow spectral line width such that the natural emission spectrum constitutes at most a few percent of the transmitted bit rate. Otherwise, the transmitter laser's emission spectrum masks the desired signal spectrum.

Semiconductor lasers, in particular, suffer from large spectral line width and frequency instability; they usually oscillate on several longitudinal modes within the laser cavity. For example, a semiconductor laser diode in a fiber-optic direct-detection system may have a line width of 400 GHz. Since typical data rates may range from 10 to 1000 Mbps for optical intersatellite links (ISL's), optical sources with line widths on the order of 100 kHz to 10 MHz are needed for coherent systems. Semiconductor lasers can achieve such bandwidths when induced to oscillate in a single longitudinal mode through the use of distributed feedback techniques or by extending the optical cavity length through the use of an external reflecting cavity. Achievement of such bandwidths by either means results in a reduction in available output power from the laser (in tens, rather than hundreds, of milliwatts).

4.1 Quantum-Limited Bit-Error-Rate Performance of Heterodyne Detection

As discussed in section 2, optical heterodyning shifts the spectrum of the desired information signal from the optical carrier frequency ω_S to the IF $\omega_{IF} = \omega_S - \omega_L$, where ω_L is the optical frequency of the LO laser. (In the case of homodyning, where $\omega_L = \omega_S$, the signal spectrum is shifted directly to baseband.) At the output of the IF filter, the desired signal current is given by equations (2.11) or (2.17), which, in general, can be expressed as

$$s(t) = A(t) \cos [\omega_o t + \phi(t)] \tag{4.1}$$

where

$$A(t) = A_S(t)A_L K = 2\sqrt{P_S(t)P_L}\,\alpha e$$

$$\omega_o = \omega_{IF} = \omega_S - \omega_L$$

$$\phi(t) = \theta_{IF}(t) = \theta_S(t) - \theta_L$$

If one assumes that the LO laser is stable (i.e., that A_L, ω_L, and θ_L are constant), then, as in digital RF communications, information can be encoded by varying the amplitude (ASK), frequency (FSK), or phase (PSK) of the transmitted optical field. In practice, modulation of the laser is accomplished either externally, with electro-optic or acousto-optic techniques, or

internally, by direct modulation of a semiconductor laser's drive current. Although these are the three basic digital modulation formats, all the modulation schemes that have been developed for RF systems can, in principle, be used in optical heterodyne systems. If T_s is the symbol duration, then the general analytic expressions of the IF signal for M-ary ASK, FSK, and PSK are, respectively,

$$\begin{aligned}
&\text{ASK: } s_i(t) = A_i(t)\cos(\omega_o t + \phi) \quad 0 \le t \le T_s \quad i = 1,\ldots,M \\
&\text{FSK: } s_i(t) = A\cos(\omega_i t + \phi) \quad 0 \le t \le T_s \quad i = 1,\ldots,M \\
&\text{PSK: } s_i(t) = A\cos[\omega_o t + \phi_i(t)] \quad 0 \le t \le T_s \quad i = 1,\ldots,M
\end{aligned} \tag{4.2}$$

where $\phi_i(t) = 2\pi i/M$. Note that, because the symbols in equation (4.2) are sinusoidal waveforms, the average power per symbol is $P = A^2/2$. Therefore, the symbol energy is

$$E = \frac{A^2 T_s}{2} = \frac{\left(2\sqrt{P_L P_S}\, e\alpha\right)^2 T_s}{2} = 2P_L P_S e^2 \alpha^2 T \tag{4.3}$$

Gagliardi (ref. 4) and others have shown that, under a strong LO condition, the LO shot noise, which becomes the dominant noise source at the detector output, can be accurately modeled as an additive, white gaussian noise random process $n_S(t)$ whose two-sided power spectral density is, from equation (2.15), written as

$$\frac{N_s}{2} = \alpha e^2 P_L \tag{4.4}$$

If we assume that the IF filter transfer function is flat (i.e., that $|H(\omega)| = 1$) over the desired IF signal bandwidth, then the received signal at the filter output is given by

$$r(t) = s_i(t) + n_s(t) \quad 0 \le t \le T_s \quad i = 1,\ldots,M \tag{4.5}$$

The function of the IF detector, therefore, is to decide which of the signals in the signal set $s_i(t)$, was transmitted, given the received signal $r(t)$. Since the LO shot noise can be modeled as zero-mean additive, white gaussian noise, the determination of the optimum-detection method and BER performance proceeds in exactly the same manner as for RF systems corrupted by zero-mean additive, white gaussian thermal noise with a two-sided spectral density of $N_o/2$. In particular, the optimum IF detector is a correlation receiver that, over each symbol period T_s, computes the correlation of the received signal $r(t)$ with each of the possible transmitted signals $s_i(t)$, and chooses the one with the largest correlation.

Since the IF receiver requires knowledge of the IF carrier phase to generate the local reference signals, the IF detection in this case is coherent. Alternatively, as with RF systems, we can use suboptimum noncoherent techniques, such as envelope detection or differential detection. Noncoherent detection is simpler and provides greater immunity to phase noise because the receiver does not require knowledge of the IF carrier phase. For this reason, noncoherent FSK (NCFSK) is often the preferred modulation method for optical heterodyne systems that use semiconductor lasers (ref. 9).

Since LO shot noise can be approximated as zero-mean, additive, white gaussian noise, the BER performance of various modulation schemes using coherent and noncoherent IF detection can be arrived at directly from the corresponding RF BER equations simply by replacing E_b/N_o with

$$\frac{E}{N_s} = \frac{2P_L P_s \alpha^2 e^2 T_s}{2\alpha e^2 P_L} = P_s \alpha T_s = K_S \text{ counts/symbol} \tag{4.6}$$

where K_S is the average number of photoelectrons per symbol, or photoelectrons per bit when $M = 2$. Table 4.1 lists the theoretical receiver sensitivities for some common binary modulation schemes. If a bipolar signal of amplitude A is assumed for the binary baseband waveform, homodyne detection provides receiver sensitivity that is 3 dB greater than that provided by the corresponding heterodyne detection, since the symbol energy is twice as large ($E = A^2 T_s$). Therefore, as indicated in table 4.1, homodyne binary phase shift keying (BPSK) has the highest sensitivity of all unencoded optical detection schemes. However, this performance improvement

TABLE 4.1.—QUANTUM-LIMITED BER PERFORMANCE OF BINARY MODULATION SCHEMES USING OPTICAL HETERODYNE AND HOMODYNE DETECTION

[Complementary error function = $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{(-u^2/2)} du$.]

Modulation	Bit error rate, BER	Modulation	Bit error rate, BER
Heterodyne detection		Homodyne detection	
ASK	$Q\left[\sqrt{K_S/2}\right]$	PSK	$Q\left[\sqrt{4K_S}\right]$
FSK	$Q\left[\sqrt{K_S}\right]$	ASK	$Q\left[\sqrt{K_S}\right]$
PSK	$Q\left[\sqrt{2K_S}\right]$		
NC FSK	$\frac{1}{2} e^{(-K_S/2)}$		
NC DPSK	$\frac{1}{2} e^{(-K_S)}$		

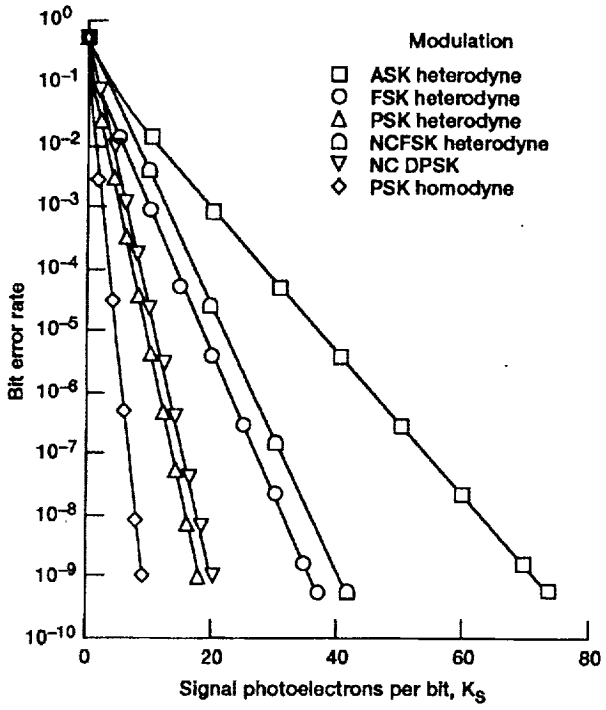


Figure 4.1.—Bit error rate (BER) versus signal photocounts/bit for heterodyne detection.

is possible only if the LO laser is phase locked to the received optical carrier through an optical phase-locked loop (PLL). Figure 4.1 plots BER versus signal counts per bit for the various modulation schemes.

4.2 Degradation Resulting From Laser Phase Noise

The BER equations given in table 4.1 assume that both the transmitting and LO laser sources are ideal, single-frequency sources with zero phase noise (i.e., zero line width). In reality, the lasers have nonzero line width, and the resulting phase noise, or jitter, on the laser carrier causes the phase of the received IF signal to fluctuate randomly in time. In the presence of this phase noise, a coherent demodulator that requires a local IF carrier that is phase and frequency locked to the unmodulated, received IF carrier will suffer BER degradation. Noncoherent demodulators, however, will be less sensitive to phase noise, since a synchronized local carrier is not needed for detection.

As will be shown, BER degradation resulting from laser phase noise is primarily a function of the laser-line-width to bit-rate ratio—the smaller the line width in comparison to the bit rate, the smaller the degradation. The following discussion focuses on semiconductor lasers, since these typically have much larger line widths than do gaseous (e.g., CO_2) or solid-state (e.g., Nd:YAG) lasers.

The spectral line shape of the unmodulated output of a semiconductor laser can be approximated by a Lorentzian function, plotted in figure 4.2 and defined by

$$L(f) = \frac{\Delta f/2\pi}{(f - f_o)^2 + (\Delta f/2)^2} \tag{4.7}$$

where f_o is the nominal center frequency and where Δf, the line width in hertz, is the full width of the function defined by the points on the curve that are at half the maximum value (i.e., the full width at half maximum). Assuming negligible amplitude noise and ignoring long-term frequency drift, the output of the laser may be modeled as

$$r(t) = A \cos [\omega_o t + \phi(t)] \tag{4.8}$$

where $\omega_o = 2\pi f_o$ is the center frequency in radians per second and where $\phi(t)$ is the random phase noise. Since instantaneous frequency is the time derivative of instantaneous phase, the frequency fluctuation noise is simply the first derivative, $\dot{\phi}(t)$.

The power spectrum of the laser frequency noise consists of three major components (ref. 9). The first component is a low-frequency disturbance called frequency flicker noise, which has a $1/f$ dependence, where f refers to frequency relative to the nominal center frequency. Frequency flicker noise is caused primarily by laser-temperature fluctuations. The second and most important component is white-frequency noise, which extends over a broad spectral range and has a one-sided, flat spectral density given by

$$S_\omega(f) = 4\pi\Delta f \qquad (\text{rad/sec})^2/\text{Hz} \tag{4.9}$$

The third component is a peak at the laser cavity's resonant frequency and typically occurs at several gigahertz. Since a well-designed system can track out the low-frequency flicker noise and since the resonant peak usually is far above the system bandwidth, only the white frequency noise component contributes to the IF signal phase noise. Recalling the Fourier transform property for time differentiation

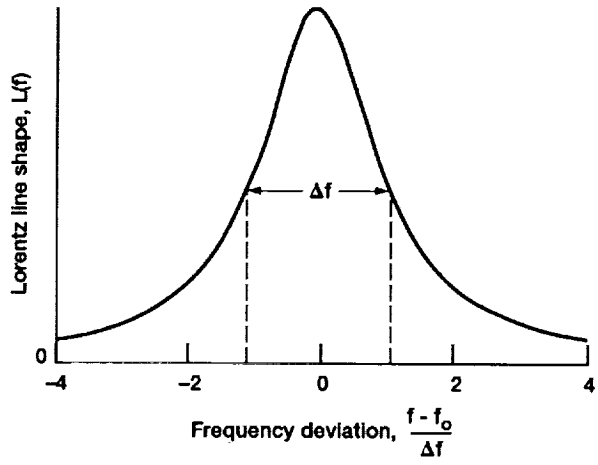


Figure 4.2.—The Lorentzian line shape function.

$$\frac{d}{dt}x(t) \Leftrightarrow (j2\pi f)X(f) \tag{4.10}$$

it is apparent that the frequency noise power spectral density (PSD) is related to the phase noise PSD $S_\phi(f)$ by

$$S_\omega(f) = (2\pi f)^2 S_\phi(f) \tag{4.11}$$

Hence, following equation (4.9), the one-sided phase noise PSD is given by

$$S_\phi(f) = \frac{S_\omega(f)}{(2\pi f)^2} = \frac{4\pi\Delta f}{(2\pi f)^2} = \frac{\Delta f}{\pi f^2} \quad \text{rad}^2/\text{Hz} \tag{4.12}$$

Because heterodyning merely translates any noise or signal spectrum from the optical carrier frequency f_o to the IF carrier frequency f_{IF}, the spectral density of the unmodulated IF signal also is Lorentzian. Because the IF signal phase noise equals the difference between two statistically independent phase noises (the transmitting laser phase noise and the LO laser phase noise), the line width of the IF signal is the sum of the individual laser line widths. Therefore, if we assume each laser has line width Δf, the one-sided PSD of the IF signal phase noise is

$$S_{\phi_{IF}}(f) = \frac{2\Delta f}{\pi f^2} \tag{4.13}$$

The PSD of the IF phase noise (eq. (4.13)) and LO shot noise (eq. (4.4)) now can be used to determine the BER degradation of heterodyne receivers that use coherent IF demodulation.

In a coherent demodulator, a local reference signal that is synchronized in both phase and frequency to the received, unmodulated carrier usually is generated by means of a phase-locked loop (PLL). A block diagram of a general-order PLL appears in figure 4.3. The general-order PLL is the simplest configuration and can only be used when the input signal

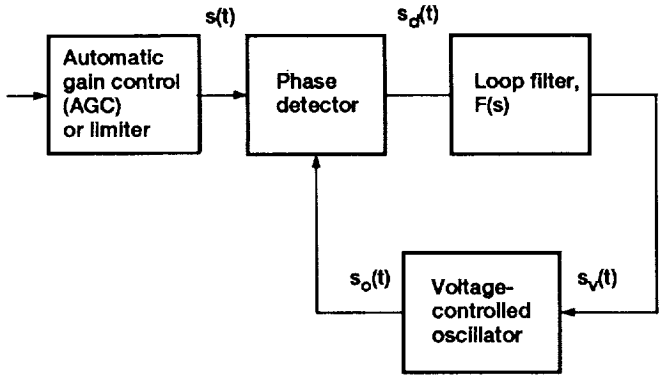


Figure 4.3.—Block diagram of a simple phase-locked loop. (See equations (4-14) to (4-19) for definitions of signals.)

spectrum has a component at the carrier frequency (e.g., when a pilot signal is transmitted along with the information signal). Other PLL configurations (e.g., a squaring PLL or Costas PLL) must be used for systems that do not transmit a pilot signal and that use modulation schemes such as binary (BPSK) or quaternary phase shift keying (QPSK), which do not contain a carrier component. The basic PLL consists of a phase detector, a loop filter with transfer function $F(s)$ and a voltage-controlled oscillator (VCO) whose output frequency deviation is proportional to its input (fig. 4.3). The output of the phase detector is proportional to the phase difference $\psi = \phi - \theta$ between the input signal and VCO output signal. The phase detector usually is modeled as an ideal multiplier followed by a low-pass filter. If we let the PLL input signal be

$$s(t) = A\cos(\omega t + \phi) \tag{4.14}$$

and the output signal from the VCO be

$$s_o(t) = A_v \cos\left(\omega t + \theta + \frac{\pi}{2}\right) = -A_v \sin(\omega t + \theta) \tag{4.15}$$

then the phase detector output before low-pass filtering is

$$\begin{aligned} s_d(t) &= s(t)s_o(t) \\ &= -AA_v \cos(\omega t + \phi)\sin(\omega t + \theta) \\ &= \frac{-AA_v}{2}[\sin(2\omega t + \phi + \theta) - \sin(\phi - \theta)] \end{aligned} \tag{4.16}$$

After low-pass filtering to remove the double frequency term, the phase-detector output is

$$s_d(t) = \frac{AA_v}{2}\sin(\phi - \theta) = K_d \sin\psi \tag{4.17}$$

Note that, because of the sine function, the phase detector has a nonlinear characteristic, and therefore the PLL represents a nonlinear system. If, however, we assume that the loop previously has acquired the input signal carrier frequency and phase and is in a tracking mode, then the phase error is small enough to allow

$$s_d(t) = K_d \sin\psi \approx K_d\psi = K_d(\phi - \theta) \tag{4.18}$$

When equation (4.18) is satisfied, the loop operates in a linear mode, and we can apply linear-control system techniques to determine the PLL closed-loop transfer function. Since the VCO generates a sinusoidal output whose frequency deviation from the input frequency is proportional to its input, the transfer characteristic of the VCO is

$$\frac{d\theta}{dt} = K_v s_v(t) \tag{4.19}$$

where K_v is a constant. If we now use the Laplace transform and take $\mathcal{L}[\phi(t)] = \phi(s)$ as the input signal and $\mathcal{L}[\theta(t)] = \Theta(s)$ as the output signal, we have, from equation (4.19),

$$s\Theta(s) = K_v S_v(s) \rightarrow \frac{\Theta(s)}{S_v(s)} = \frac{K_v}{s} \tag{4.20}$$

as the transfer function of the VCO. The open-loop transfer function is then simply

$$G(s) = \frac{K_d K_v F(s)}{s} \tag{4.21}$$

Since unity feedback is present, the closed-loop transfer function is

$$H(s) = \frac{\Theta(s)}{\Phi(s)} = \frac{G(s)}{1 + G(s)} = \frac{K_d K_v F(s)/s}{1 + K_d K_v F(s)/s} = \frac{K_d K_v F(s)}{s + K_d K_v F(s)} \tag{4.22}$$

Finally, the phase-error transfer function is given by

$$\frac{\Psi(s)}{\Phi(s)} = \frac{\Phi(s) - \Theta(s)}{\Phi(s)} = 1 - H(s) \tag{4.23}$$

From equation (4.22), we see that the loop dynamics (i.e., zeros and poles) are controlled by the loop filter transfer function $F(s)$. When equation (4.22) is of the second order, it is customary to express its denominator in the form

$$s^2 + 2\varsigma\omega_n s + \omega_n^2 \tag{4.24}$$

where ς and ω_n are the damping factor and natural frequency, respectively. Two other important parameters of the PLL are the open-loop dc gain

$$K = K_d K_v F(0) \tag{4.25}$$

and the noise-equivalent single-sided bandwidth

$$B_N = \frac{1}{H_o^2} \int_0^\infty |H(f)|^2 df \tag{4.26}$$

In equation (4.26), $H(f)$ is obtained from $H(s)$ by making the substitution $s = j2\pi f$, and H_o is the maximum gain (i.e., $H_o = |H(f)|_{\max}$). Recall that the noise equivalent bandwidth is so named because it represents the bandwidth of an ideal rectangular filter with constant gain H_o, which passes the same amount of noise power as does the actual filter with transfer function $H(f)$ with white noise present at its input.

Accordingly, if the PLL is operating in the linear (i.e., tracking) mode and the input is the IF signal (with phase noise) plus LO shot noise (modeled as additive, white gaussian noise), then the phase error $\psi(t)$ can be approximated as a zero-mean, gaussian random process whose variance is of the form (ref. 10)

$$\sigma_\psi^2 = \sigma_{SN}^2 + \sigma_{PN}^2 \tag{4.27}$$

Here, σ_{SN}^2 is the phase error variance resulting from LO shot noise, and σ_{PN}^2 is the phase-error variance resulting from laser phase noise. Because linear PLL operation is assumed, the shot-noise contribution is analogous to that which occurs for additive, white gaussian thermal noise and is given by

$$\sigma_{SN}^2 = \frac{2N_s B_N}{A_{IF}^2} \tag{4.28}$$

where N_s is the LO shot-noise single-sided PSD, and A_{IF} is the IF carrier amplitude. Using the IF signal phase-noise single-sided PSD in equation (4.13) and the phase error transfer function in equation (4.23) gives a phase noise contribution to the phase-error variance of

$$\sigma_{PN}^2 = \int_0^\infty S_{\psi_{IF}}(f)\,|1 - H(f)|^2 df = \frac{2\Delta f}{\pi} \int_0^\infty \frac{|1 - H(f)|^2}{f^2} df \tag{4.29}$$

If, for simplicity, we approximate $H(f)$ by an ideal low-pass filter of bandwidth B_N, then equation (4.29) can be simplified to

$$\sigma_{PN}^2 = \frac{2\Delta f}{\pi B_N} \tag{4.30}$$

Of course, a more accurate value requires specific knowledge of the PLL closed-loop transfer function $H(f)$. Spilker (ref. 11) describes a particular, widely used, second-order PLL for which the damping factor is $\varsigma = 1/\sqrt{2}$ (for good transient response) and the phase-error variance is

$$\sigma_{PN}^2 = 3.7 \frac{2\Delta f}{\pi B_N} \tag{4.31}$$

In general, other loop-transfer functions will yield different values for the multiplying constant. If we let the multiplying constant be denoted by ρ and make the appropriate substitutions for

N_s and A_{IF}, the total phase-error variance can be expressed as

$$\sigma_\psi^2 = \frac{2N_sB_N}{A_{IF}^2} + \rho\,\frac{2\Delta f}{\pi B_N}$$

$$= \frac{2(2\alpha e^2 P_L)B_N}{\left(2\sqrt{P_L P_s}\,\alpha e\right)^2} + \rho\,\frac{2\Delta f}{\pi B_N} \qquad (4.32)$$

$$= \frac{B_N}{\alpha P_s}\,\frac{T}{T} + \rho\,\frac{2\Delta f}{\pi B_N} = \frac{B_N T}{K_S} + \rho\,\frac{2\Delta f}{\pi B_N}$$

Note from equation (4.32) that decreasing the loop bandwidth decreases the phase-error variance resulting from LO shot noise but increases the error variance resulting from laser phase noise. Since we would like to minimize the total phase-error variance, the optimum value of the loop bandwidth can be found by taking the first derivative of equation (4.32) with respect to B_N, setting it equal to zero, and solving for B_N. Hence,

$$\frac{\partial \sigma_\psi^2}{\partial B_N} = \frac{T}{K_s} - \frac{2\Delta f\rho}{\pi B_N^2} = 0 \qquad (4.33)$$

$$B_{N_{opt}} = \sqrt{\frac{2\rho K_S \Delta f}{\pi T}} = \sqrt{\frac{2\rho K_S \Delta f T}{\pi}}\,R_s \qquad (4.34)$$

where $R_S = 1/T$ is the symbol rate. By substituting $B_{N_{opt}}$ for B_N in equation (4.32), we arrive at the minimum phase-error variance in terms of the laser line width-to-symbol rate ratio of

$$\sigma_{\psi_{min}}^2 = \sqrt{\frac{2\rho\Delta f/R_s}{\pi K_s}} + \frac{2\rho\Delta f/R_s}{\sqrt{2\rho K_S \pi \Delta f/R_s}} \qquad (4.35)$$

Now consider the effect of zero-mean gaussian random phase error with a variance equal to equation (4.35) on the BER of those systems using IF coherent demodulation. Suppose that the received signal without noise in a BPSK system during a particular symbol interval is of the form

$$s(t) = A\cos(\omega_{IF}t + \phi_o) \qquad (4.36)$$

If we demodulate the signal by multiplying $s(t)$ with a reference IF carrier $c(t)$ whose phase is not ϕ_o but $\hat{\phi}_o$, we obtain

$$s(t)c(t) = A\cos(\omega_{IF}t + \phi_o)\cos(\omega_{IF}t + \hat{\phi}_o)$$

$$= \frac{A}{2}\left[\cos(2\omega_{IF}t + \phi_o + \hat{\phi}_o) + \cos(\phi_o - \hat{\phi}_o)\right] \qquad (4.37)$$

which, after low-pass filtering, becomes

$$y(t) = \frac{A}{2}\cos(\phi_o - \hat{\phi}_o) = \frac{A}{2}\cos(\psi) \qquad (4.38)$$

Thus, the effect of a static phase error ψ is to reduce the signal amplitude by the amount of cos ψ and therefore the signal power and symbol energy by the amount of $\cos^2\psi$. The same result holds for coherent binary FSK and ASK. The effect of carrier phase error is even more severe on multiphase PSK and quadrature (QAM) since, in addition to causing the power reduction, the phase error results in cross-talk interference between the signal's in-phase and quadrature components. Therefore, under static phase-error conditions, the BER's of the coherent schemes listed in table 4.1 are found by multiplying K_S in the expressions by $\cos^2\psi$. For example, for BPSK, we would have

$$BER = Q\left[\sqrt{2K_S\cos^2\psi}\right] \qquad (4.39)$$

Finally, under a random phase error condition wherein the phase error is modeled by a gaussian probability density, the average BER can be found by taking the expected value of equation (4.39). As such, in the presence of laser phase noise, the average BER of heterodyne BPSK is given by

$$BER_\psi = E\left[Q\left[\sqrt{2K_S\cos^2\psi}\right]\right]$$

$$= \frac{1}{\sqrt{2\pi}\sigma_\psi}\int_{-\infty}^{\infty} Q\left[\sqrt{2K_S\cos^2\psi}\right] e^{(-\psi^2/2\sigma_\psi^2)}d\psi \qquad (4.40)$$

where σ_ψ^2 is the phase-error variance given in equation (4.32). Similarly, the average BER of heterodyne BFSK is given by

$$BER_\psi = \frac{1}{\sqrt{2\pi}\sigma_\psi}\int_{-\infty}^{\infty} Q\left[\sqrt{K_S\cos^2\psi}\right] e^{(-\psi^2/2\sigma_\psi^2)}d\psi \qquad (4.41)$$

and the average BER for coherent binary ASK, by

$$BER_\psi = \frac{1}{\sqrt{2\pi}\sigma_\psi}\int_{-\infty}^{\infty} Q\left[\sqrt{\frac{K_S\cos^2\psi}{2}}\right] e^{(-\psi^2/2\sigma_\psi^2)}d\psi \qquad (4.42)$$

The degradation resulting from laser line width can be evaluated by use of these equations. For example, table 4.2 lists the power penalty associated with the laser phase noise for a heterodyne BPSK system for several values of laser line width-to-bit rate ratio at 10^{-6} and 10^{-9} BER. (The power penalty is defined here as the increase in the number of signal photoelectrons per bit K_S, in decibels, necessary to maintain

TABLE 4.2.—POWER PENALTY RESULTING FROM LASER PHASE NOISE FOR HETERODYNE BPSK USING OPTIMAL LOOP BANDWIDTH AND $\rho = 3.7$

Laser-line-width to bit-rate ratio, $\Delta f/R_b$, percent	Bit error rate, BER			
	10^{-6}		10^{-9}	
	$K_{S_{ideal}}$			
	12		18	
	K_S	Loss, dB	K_S	Loss, dB
0.5	14	0.7	25	1.4
1.0	17	1.5	35	2.9
1.5	21	2.4	45	4.0
2.0	26	3.4	55	4.9
2.5	30	4.0	67	5.7
5.0	55	6.6	130	8.6
10.0	110	9.6	259	11.6

a given BER in relation to the quantum-limited value found from the expressions in table 4.1.) The values in table 4.2 were computed with the optimum PLL bandwidth in equation (4.35) to minimize phase-error variance with $\rho = 3.7$. Note that even with the optimum loop bandwidth, laser line width must be held to about 1 percent of the bit rate to avoid significant power penalty.

Laser phase noise also causes degradation of noncoherent detection schemes, such as envelope-detected orthogonal FSK and differentially detected differential phase shift keying (DPSK). Phase noise degrades the orthogonality of the FSK tones so that in addition to being a function of the photoelectrons per bit, the BER becomes a function of the tone spacing and laser line width. Owing to the complexity of the BER derivation of noncoherently detected orthogonal FSK in the presence of phase noise, only the final results from reference 9 will be given here. If we assume that the phase noise is modeled as a random frequency noise γ that is zero-mean gaussian with a variance of $\sigma_\gamma^2 = \Delta f R_b/\pi$, the BER of noncoherent binary FSK can be approximated by

$$BER = \frac{1}{\sqrt{2\pi}\sigma_\gamma}\int_{-\infty}^{\infty}\frac{1}{2}\left[1 - Q_M\left(\sqrt{K_S}, \sqrt{\beta(\gamma)}\right) + Q_M\left(\sqrt{\beta(\gamma)}, \sqrt{K_S}\right)\right] e^{(-\gamma^2/2\sigma_\gamma^2)}d\gamma \tag{4.43}$$

where

$$\beta(\gamma) = K_S\left[\frac{\sin(\pi(\gamma + \nu_d)T)}{\pi(\gamma + \nu_d)T}\right]^2$$ and ν_d is the tone spacing

The Marcum Q function $Q_M(a,b)$ is defined by

$$Q_M(a,b) = \int_b^\infty e^{[-(a^2+x^2)/2]} I_o(ax)x\,dx \tag{4.44}$$

TABLE 4.3.—POWER PENALTY RESULTING FROM LASER PHASE NOISE FOR NONCOHERENT BFSK

Laser-line-width to bit-rate ratio, $\Delta f/R_b$, percent	Bit error rate, BER			
	10^{-6}		10^{-9}	
	Loss, dB			
	$\nu_d = R_b$	$\nu_d = 2R_b$	$\nu_d = R_b$	$\nu_d = 2R_b$
0.5	0.2	⁻0	0.3	⁻0
1.0	.5	.1	1.1	.2
1.5	.8	.2	3.0	.5
2.0	1.7	.4	9.1	.8
2.5	3.0	.5	(a)	1.3
5.0	(a)	2.1	(a)	6.6
10.0	(a)	12.5	(a)	(a)

[a]For these entries no amount of signal power can achieve the desired BER since it is below the irreducible error rate.

and $I_o(z)$ is the modified Bessel function of the first kind of order zero given by

$$I_o(z) = \frac{2}{\pi}\int_0^{\pi/2}\cosh(z\sin\theta)\,d\theta \tag{4.45}$$

Table 4.3 lists the power penalties derived through these equations for two values of the tone spacing, at once and twice the bit rate, respectively, and for several values of the laser line width-to-bit rate ratio. As we would expect, these values indicate that noncoherent BFSK is less sensitive to the phase noise than is coherent BPSK and that increasing the tone spacing for a given line width reduces the sensitivity. Because the phase noise is not additive noise, which can be overcome by simply increasing the signal power, the phase noise for some of the entries in table 4.3 attains such a level of dominance that the desired BER cannot be achieved regardless of the magnitude of K_S. The irreducible error rate, or error-rate floor, introduced by the phase noise has been estimated (ref. 9) to be

$$BER_{\text{floor}} = Q\left[\frac{\nu_d/R_b}{\sqrt{4\Delta f/(R_b\pi)}}\right] NFSK \tag{4.46}$$

Therefore, for a given bit rate, increasing the tone spacing lowers the error-rate floor, whereas increasing the line width raises the floor.

Like that for noncoherent FSK, an exact BER analysis for DPSK in the presence of phase noise is extremely difficult to obtain. However, for small line-width to bit-rate ratios, the phase noise corrupted BER for DPSK can be approximated (ref. 9) by

$$BER = \frac{1}{2}e^{-K_S} + Q\left[\frac{\sqrt{2}\pi/2}{\sqrt{2/K_S + 16\pi\Delta f/(3R_b)}}\right] \tag{4.47}$$

Taking the limit of equation (4.47) as $K_S \to \infty$, the BER floor is

$$BER_{\text{floor}} = Q\left[\sqrt{\frac{3\pi}{32\Delta f/R_b}}\right] \quad \text{DPSK} \tag{4.48}$$

When we use equation (4.48), the error-rate floor becomes less than 10^{-9} once the laser line width has reached 1 percent of the data rate, and less than 10^{-6} once the laser linewidth has 1.5 percent of the data rate.

5.0 Determination of Interfering Background Power

In addition to collecting desired-signal power, the optical receiver also collects interfering background radiation that degrades system performance. Section 5.1 presents equations for computing background power in which the background source is modeled as a black-body radiator. Section 5.2 discusses the relative immunity of the heterodyne optical receiver to background radiation.

5.1 Computation of Background Power

Recall that a plane wave arriving normal to the receiving aperture plane produces a Fraunhofer diffraction, or Airy pattern, in the telescope's focal plane, with the angular radius of the Airy disk, or focal spot, being given by $\theta_o = 1.22\lambda/D$, where D is the aperture diameter. Here, we are assuming the presence of diffraction-limited optics (i.e., image quality is limited only by diffraction effects because lens and mirror aberrations are negligible). Plane waves that arrive at the receiver off-axis produce Airy patterns in the focal plane that are displaced from the detector's center. If a plane wave arrives at the receiver at a large enough off-axis angle, the plane wave's pattern shifts off the detector's surface and cannot even be detected. Therefore, the receiver field of view (FOV) can be defined as the solid angle, looking out from the detector surface, within which all plane waves must arrive in order to project their diffraction patterns on the detector. A conical FOV, described by a planar angle θ (full angle in radians) is related to the corresponding solid angle FOV Ω (in steradians) by

$$\Omega = 4\pi \sin^2\frac{\theta}{4} \quad \text{sr} \tag{5.1}$$

If we assume a circular detector of radius r_d and area $A_d = \pi r_d^2$, then the receiver planar-angle FOV is

$$\theta_{fv} = 2\tan^{-1}\frac{r_d}{f} \approx \frac{2r_d}{f} \tag{5.2}$$

and the receiver solid-angle FOV is

$$\Omega_{fv} = 4\pi \sin^2\frac{\theta_{fv}}{4} \approx 4\pi\left(\frac{\theta_{fv}}{4}\right)^2 = \frac{\pi r_d^2}{f^2} = \frac{A_d}{f^2} \tag{5.3}$$

where f is the focal length. In equations (5.2) and (5.3), the small-angle approximation, $\sin\theta \approx \tan\theta \approx \theta$ for $\theta \le 0.1$ rad, has been used. Since the measure of a telescope's resolution is given by the Rayleigh criterion, two distant point sources, each of which produces an Airy pattern image in the focal plane, are "just resolvable" if their angular separation equals the angular radius of the Airy disc of either image ($\theta_o = 1.22\lambda/D$). If their angular separation is less than this value, the two images overlap too much to be distinguished as separate sources and, thus, are termed "unresolvable." Although this measure of resolution is somewhat arbitrary, it enables the receiver diffraction-limited FOV to be defined as the solid angle, looking out from the detector, within which all plane waves are indistinguishable in terms of their direction of arrival. Thus, the images of plane waves arriving within the diffraction-limited FOV are superimposed on one another and appear as a single source. Following equation (5.2), the diffraction-limited planar angle FOV is $\theta_{dL} = 2\theta_o$, and the diffraction-limited solid angle FOV is

$$\Omega_{dL} = 4\pi \sin^2\left(\frac{2\theta_o}{4}\right) \approx \pi\theta_o^2 = (1.22)^2\pi\frac{\lambda^2}{D^2} = \frac{\pi^2(1.22)^2}{4}\frac{\lambda^2}{A_r} \tag{5.4}$$

where A_r is the receiver aperture area. Figure (5.1) shows the relationship between the receiver FOV and its diffraction-limited FOV. Note from equations (5.3) and (5.4) that, since λ is on the order of microns, whereas detector dimensions are on the order of millimeters in an optical system, the receiver FOV typically is much larger than the diffraction-limited FOV. For example, if the focal length f is set to approximately the aperture diameter D, the ratio of receiver FOV to diffraction-limited FOV is

$$\frac{\Omega_{fv}}{\Omega_{dL}} = \frac{1}{(1.22)^2}\frac{r_d^2}{\lambda^2} \tag{5.5}$$

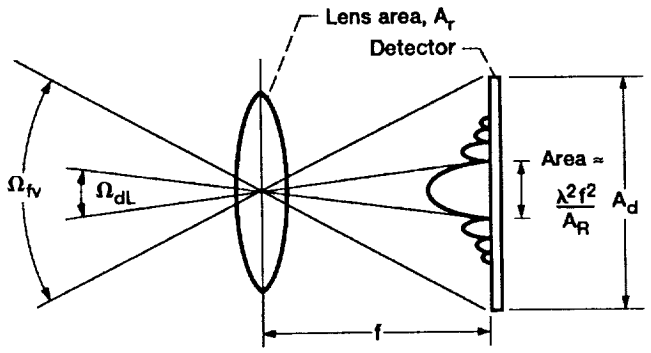


Figure 5.1.—Relationship between receiver field of view, Ω_{fv}, and diffraction-limited field of view, Ω_{dL}.

Again as an example, for a 1-μm wavelength and a 1-mm diameter detector, this ratio is about 100 000. The ratio Ω_{fv}/Ω_{dL} is considered to be the optical field's number of spatial modes in that waves occurring within a single spatial mode or diffraction-limited FOV cannot be resolved or differentiated from one another (ref. 4).

Since many sources of background radiation, such as stars, have spectral characteristics approximating those of a black body, Planck's formula for black body spectra can be used to estimate the amount of collected background power. For a black body of temperature T in degrees kelvin, the spectral radiance—the power radiated at frequency ν, per unit of source area, into a unit solid-angle, per-unit bandwidth—is given by

$$I(\nu,T) = \frac{2h\nu^3/c^2}{e^{(h\nu/kT)} - 1} \qquad \text{W/(m}^2 - \text{Hz} - \text{sr)} \qquad (5.6)$$

where

h Planck constant, 6.626×10^{-34} J-sec

k Boltzmann constant, 1.381×10^{-23} J/K

c light speed in vacuum, 2.998×10^{8} m/sec

The Planck function also can be expressed in terms of wavelength by setting the power emitted by the source per unit-of-area per unit-of-solid-angle in a frequency interval $d\nu$ to the power emitted per unit-of-area per unit-of-solid-angle in a wavelength interval $d\lambda$ as

$$I(\nu,T)\,d\nu = I(\lambda,T)\,d\lambda \qquad (5.7)$$

Since $\nu = c/\lambda$ and $d\nu/d\lambda = -c/\lambda^2$ (the minus sign indicates that frequency decreases as wavelength increases), we have

$$I(\lambda,T) = I(\nu,T)\frac{d\nu}{d\lambda} = I(\nu,T)\frac{c}{\lambda^2}$$

$$= \frac{2hc^2/\lambda^5}{e^{(hc/\lambda kT)} - 1} \qquad \text{W/(m}^2 - \text{m} - \text{sr)} \qquad (5.8)$$

Table 5.1 lists the spectral radiance of various bodies at two wavelengths as computed from equation (5.8).

Now, consider a receiving aperture of area A_r located at distance Z from a background source of circular disk area A_s. Following equation (5.3), the solid angle subtended by the source when viewed from the receiver is approximately $\Omega_s = A_s/Z^2$ sr. Similarly, the solid angle subtended by the receiver aperture when viewed from the source is $\Omega_r = A_r/Z^2$ sr. We will examine two types of background radiators: extended background sources, such as sky backgrounds, which occupy the entire receiver FOV (i.e., $\Omega_s \geq \Omega_{fv}$), and discrete or localized sources, such as the Sun, planets, or stars, which occupy only a fraction of the receiver FOV (i.e., $\Omega_s < \Omega_{fv}$). Using equation (5.6) and the preceding results yields the background power collected by the receiver in a bandwidth of B hertz, about the

TABLE 5.1.—CALCULATED SPECTRAL RADIANCE OF VARIOUS BODIES USING PLANCK'S BLACK-BODY RADIATION EQUATION

	Approximate black-body temperature, K	Spectral radiance, W/cm^2-sr-μm	
		Wavelength, λ, μm	
		1	10
Sun	6000	1192	0.44
Moon	373	2.13×10^{-13}	0.003
Mercury	613	7.68×10^{-7}	0.013
Venus	235	3.12×10^{-23}	2.62×10^{-4}
Earth	300	1.79×10^{-17}	9.94×10^{-4}
Mars	217	1.94×10^{-25}	1.58×10^{-4}
Jupiter	138	6.44×10^{-42}	3.54×10^{-6}
Saturn	123	1.94×10^{-47}	9.94×10^{-7}
Uranus	90	4.64×10^{-66}	1.36×10^{-8}

frequency ν, of

$$\left.\begin{aligned} P_B &= I(\nu,T)A_sB\Omega_r\left(\frac{\Omega_{fv}}{\Omega_s}\right) \quad \text{(extended source)} \\ P_B &= I(\nu,T)A_sB\Omega_r \qquad\qquad\quad \text{(local source)} \end{aligned}\right\} \qquad (5.9)$$

Since $A_s = \Omega_s Z^2$ and $\Omega_r = A_r/Z^2$, equation (5.9) can be rewritten as

$$P_B = I(\nu,T)BA_r\Omega_{fv} \quad \text{(extended)} \qquad (5.10a)$$

$$P_B = I(\nu,T)BA_r\Omega_s \quad \text{(local)} \qquad (5.10b)$$

Thus, for an extended background source that fills the entire receiver FOV, the interfering power depends on the receiver aperture area and the receiver FOV. For a local source that appears as a spot in the receiver FOV, the collected power is a function of the aperture area and the solid angle subtended by the source at the receiver. Usually, for point sources viewed from Earth, background radiation is given in terms of spectral irradiance $H(\nu)$ rather than in terms of the spectral radiance $I(\nu)$. The two are related by

$$H(\nu) = \Omega_s I(\nu) \qquad \text{W/(m}^2 - \text{Hz)} \qquad (5.11)$$

Therefore, the spectral irradiance is the power received per unit bandwidth per unit receiver area. Measurements of the spectral irradiance of background sources such as the Moon, planets, and bright stars, are easily obtained in physics and optics literature. For a given value of spectral irradiance, the collected background power is

$$P_B = H(\nu)BA_r \qquad (5.12)$$

5.2 Immunity of Heterodyne Receiver to Background Radiation

Using the equations in section 5.1, we can show that a heterodyne or homodyne optical receiver is virtually immune to the noncoherent radiation of thermal background sources, at least for wavelengths below 1μm. This immunity results from the heterodyne detection's photomixing process, which provides both spectral and spatial discrimination of background radiation. Spectral discrimination occurs because only background noise components falling within the relatively narrow IF bandwidth are amplified and thus contribute to the IF signal noise. Spatial discrimination occurs through the same mechanism that requires the proper alignment of the received signal and LO optical fields.

Recall that heterodyning takes place where the diffraction patterns of the signal and LO fields overlap on the detector surface—the area of the Airy disk. Therefore, by definition, only background radiation arriving at the receiver within the receiver diffraction-limited FOV mixes with the LO field and generates noise components at the IF frequency. The remaining radiation, which occurs outside the diffraction-limited FOV, is focused elsewhere on the detector and simply contributes to the detector shot noise. Under normal, strong LO conditions, however, this background shot noise is insignificant compared with the LO's shot noise. Thus, background power for a heterodyne receiver is computed, only radiation occurring within the diffraction-limited FOV, Ω_{dL}, and falling within the IF passband of the post-detection filter is of concern. On the other hand, when background power for a direct-detection receiver is computed, the optical bandwidth and full receiver FOV Ω_{fi} must be considered. Because these quantities are much larger than those used to compute the background power for the heterodyne system, direct detection is much more susceptible to background radiation degradation. Front-end optical filter bandwidths are typically several orders of magnitude greater than those encountered in RF systems. Usually, optical bandwidth is expressed in wavelengths rather than in frequency. The optical filter bandwidth in wavelengths $\Delta\lambda$, about a center wavelength λ_o, is related to the optical bandwidth in hertz Δf, about a center frequency $f_o = c/\lambda_o$, by

$$\frac{\Delta\lambda}{\lambda_o} \approx \frac{\Delta f}{f_o} \tag{5.13}$$

For example, a narrow-band optical filter may have an optical bandwidth in the range of 1 to 10 Å. At 1-μm wavelength, the corresponding frequency bandwidth is 30 to 300 GHz. Also, as indicated by equation (5.5), the optical receiver FOV typically is orders of magnitude greater than the diffraction-limited FOV.

A heterodyne receiver pointed directly at the Sun with the Sun filling the entire receiver FOV illustrates the insensitivity of heterodyne detection to background noise. Over the visible and infrared wavelength region, the Sun's spectrum can be closely approximated by the spectrum of a black-body radiator at temperature $T_{sun} = 6000$ K (fig. 5.2) (ref. 12). The average received background power is found through the IF bandwidth and diffraction-limited FOV in equation (5.10a)

$$P_B = I(\nu, T_{\text{sun}}) B_{IF} A_r \Omega_{dL}$$

$$= \frac{2h\nu^3/c^2}{e^{h\nu/kT_{\text{sun}}} - 1} B_{IF} A_r \frac{\lambda^2}{A_r}$$

$$= \frac{2h\nu B_{IF}}{e^{h\nu/kT_{\text{sun}}} - 1} \tag{5.14}$$

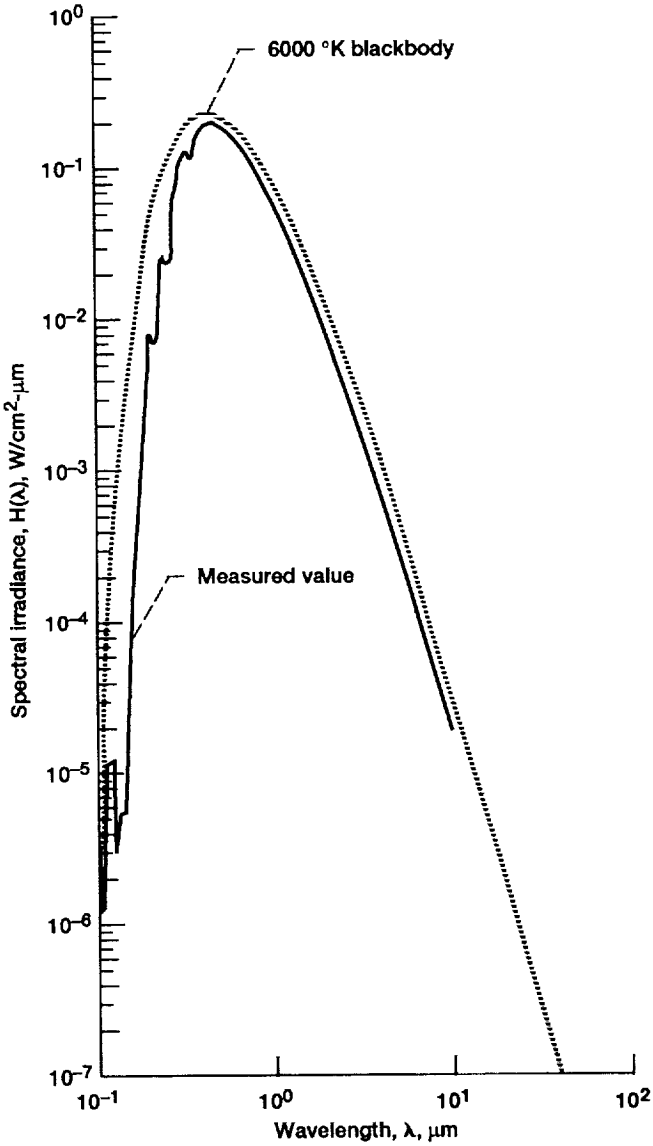


Figure 5.2.—Comparison of Sun's spectrum and spectrum of 6000-K black body.

Here, Ω_{dL} has been approximated by λ^2/A_r. (The constant factor in equation (5.4) about 3.672 can be ignored without significantly changing the end result.) In addition, on the average, only half of this radiation is of the proper polarization for detection. Therefore, the average power incident on the detector is

$$P_B = \frac{h\nu B_{IF}}{e^{(h\nu/kT_{sun})} - 1} \tag{5.15}$$

If we treat the background signal as a desired signal (i.e., P_B as P_S) and follow equation (5.16) in section 2, the quantum-limited *SNR* is

$$SNR = \frac{\eta P_S}{h\nu B_{IF}} = \eta \frac{\left(\frac{h\nu B_{IF}}{e^{(h\nu/kT_{sun})} - 1}\right)}{h\nu B_{IF}} = \frac{\eta}{e^{(h\nu/kT_{sun})} - 1} \tag{5.16}$$

For a quantum efficiency of 80 percent and a semiconductor laser-diode wavelength of 0.85 μm, the *SNR* is about 0.051 (−13 dB). Thus, if the receiver uses an aperture filter to prevent damaging heat buildup in the focal plane, the additional noise caused by the Sun coming into the receiver FOV is negligible (ref. 13). For a thermal background source to yield an *SNR* of at least one, its temperature must be at least (setting *SNR* = 1 in equation (5.16))

$$T \geq \frac{h\nu}{k} \frac{1}{\ln(\eta + 1)} \tag{5.17}$$

For the quantum efficiency and wavelength used previously, this yields a source temperature of at least 29 000 K. The quantity $(e^{(h\nu/kT)} - 1)^{-1}$ can be interpreted as the average number of photons per spatial mode emitted from a black body, or the photon degeneracy factor (ref. 5).

6.0 Optical-Link Performance in the Presence of Random Pointing and Tracking Errors

This section presents the equations used to estimate optical-link BER in the presence of random spatial pointing and tracking errors when the spatial acquisition process between the two optical terminals has already been completed by using beacon signals and an appropriate search strategy. Because of tracking sensor noise, host satellite jitter, spacecraft relative motion, telescope gimbal friction, errors in point-ahead calculation, and other errors, the instantaneous pointing and tracking errors fluctuate randomly in time and are modeled as Rayleigh-distributed random variables. Herein, we show that in the presence of random pointing and tracking errors, an optimal transmit aperture size exists for the direct detection (DD) system that minimizes the required transmitter laser power for a given average BER (ref. 9). Similarly, assuming that the LO tracking (alignment) error also is random, we show that optimum transmitter and receiver aperture sizes exist for the heterodyne and homodyne systems that minimize the required transmission power for a specified average BER. As such, we show that the link BER cannot be arbitrarily improved by increasing telescope aperture size, in contrast to the case in an ideal system characterized by zero pointing error and perfect tracking. Link optimization (i.e., designing the link for minimum transmission power requirements) for both DD M-ary PPM and heterodyne and homodyne systems is described.

A quadrant photodetector consisting of four photodetector cells matched in detection sensitivity commonly is used in spatial tracking systems to track the angular position of a remote transmitter after the initial acquisition is complete. The remote transmitter's angular position (with respect to the receiving line of sight, (LOS)) is estimated from the transmitted signal image position, or focal spot, on the quadrant photodetector, which is placed in the receiving telescope's focal plane. The current from each detector cell is used to compute estimates of the remote transmitter's angular position in the azimuthal and elevation directions. These estimates are used to generate an output signal that is proportional to the angular difference between the current receiver LOS and the transmitter's estimated position. This control signal then drives a servosystem that attempts to realign the receiver and transmitter LOS with the remote transmitter estimated position. For long-range links, a point-ahead angle also must be calculated and added to the realignment angle.

Because the quadrant detector decouples the azimuthal tracking error θ_x from the elevation tracking error θ_y, each may be modeled as a gaussian random variable with zero mean and standard deviation σ. Since the realigned receiver LOS becomes the reference direction for the new transmitter LOS, errors in tracking the remote terminal translate into transmitter pointing errors. Because the pointing and tracking error along each of the telescope gimbal axes is assumed to be a zero-mean, gaussian random process, the error variance equals the mean-square value or average power given by

$$\sigma^2 = \int_0^\infty S(f)\, df \tag{6.1}$$

where $S(f)$ is the tracking-error PSD at the spatial tracking loop output. If tracking sensor noise and platform jitter are the dominant noise sources, $S(f)$ is (from linear system theory) given by

$$S(f) = S_D(f)|H(f)|^2 + S_p(f)|1 - H(f)|^2 \tag{6.2}$$

where $S_D(f)$ and $S_p(f)$ are the single-sided PSD's of the tracking detector noise and platform jitter at the input of the spatial tracking loop, and where $H(f)$ is the closed-loop transfer function of the spatial tracking loop. By approximating $H(f)$ by an ideal low-pass filter with a single-sided noise equivalent bandwidth B_L, we can express the total tracking error variance as

$$\sigma^2 = \int_0^{B_L} S_D(f)\, df + \int_{B_L}^{\infty} S_p(f)\, df \qquad (6.3)$$

Note that as the loop bandwidth increases, the contribution from platform jitter to the rms pointing and tracking error σ decreases, whereas the contribution from tracking sensor noise increases. Since practical tracking systems have low bandwidths (from 100 Hz to 1 kHz), base motion and mechanical vibration usually are the dominant sources of pointing and tracking error, whereas the effect of tracking sensor noise is negligible. In such a situation, the rms error cannot be reduced by simply increasing the signal power to improve the tracking detector SNR. Rather, the amount of platform jitter must be reduced, and the optical platform must be isolated from the host satellite. Analysis of measurements of satellite disturbances on-board the LANDSAT spacecraft indicate that an rms tracking error of 0.5 to 1 μrad is achievable (ref. 14). By transforming the azimuth and elevation gaussian random variables, θ_x and θ_y, to polar coordinates, it can be shown that the instantaneous radial pointing and tracking error θ_t is Rayleigh distributed and has a probability density function (PDF) of

$$p(\theta_t) = \frac{\theta_t}{\sigma^2} e^{-\theta_t^2/2\sigma^2} \qquad \theta_t \geq 0 \qquad (6.4)$$

If a static pointing error is present in addition to random error (i.e., if θ_x and θ_y have mean values μ_x and μ_y), the instantaneous radial pointing error is Rician distributed with a PDF of

$$p(\theta_t) = \frac{\theta_t}{\sigma^2} e^{[-(\theta_t^2+b^2)/2\sigma^2]} I_o\left(\frac{\theta_t b}{\sigma^2}\right) \qquad (6.5)$$

where the pointing bias $b = \sqrt{\mu_x^2 + \mu_y^2}$, and $I_o(x)$ is the modified Bessel function of the first kind. As might be expected, the presence of both static and random pointing error will result in worse link performance than random pointing error alone. To simplify the analysis, we assume that there is zero pointing bias (i.e., b = 0 and the pointing error PDF is Rayleigh distributed). However, BER performance in the presence of both static and random error can be determined by substituting the Rician PDF for the Rayleigh PDF in the appropriate equations. The determination of average BER in the presence of Rayleigh-distributed pointing and tracking error is discussed for the M-ary PPM direct detection link in section 6.1 and for the heterodyne and homodyne link in section 6.2.

6.1 Pointing and Tracking Error in the M-ary Pulse-Position Direct-Detection Link

For an M-ary PPM direct-detection system operating at wavelength λ over a link range of Z and having an instantaneous pointing error of θ_t, the number of received signal photoelectrons per slot is given by

$$K_S(\theta_t, P_t, d_t) = \left[MP_t\eta_t \left(\frac{\pi d_t}{\lambda}\right)^2 \times L_t(\theta_t, d_t) \left(\frac{\lambda}{4\pi Z}\right)^2 \eta_r \left(\frac{\pi d_r}{\lambda}\right)^2 L_o \right] \frac{\eta}{h\nu}\tau \qquad (6.6)$$

where

P_t	average transmission power, W
M	PPM order
η_t, η_r	efficiency of transmitter and receiver optics, respectively
d_t, d_r	diameter of transmitter and receiver apertures, m
η	quantum efficiency of detector
$h\nu$	photon energy, hc/λ, J/photon
τ	PPM time-slot width, sec
L_o	other link losses (such as obscuration and synchronization)

and $L_t(\theta_t, d_t)$ is the pointing loss factor. For the DD link, it is assumed that the detector surface area is sufficiently large so that the remote transmitter remains within the receiver FOV (eq. (5.3)) during tracking so that there is negligible receiving pointing loss (i.e., it is only necessary that the transmitter image be focused somewhere on the detector surface). If the transmitter is approximated by a uniformly illuminated circular aperture, the intensity distribution in the far-field has an Airy diffraction pattern so that the pointing loss factor can be written as

$$L_t(\theta_t, d_t) = \left[\frac{2J_1(\pi\theta_t d_t/\lambda)}{\pi\theta_t d_t/\lambda}\right]^2 \qquad (6.7)$$

where $J_1(x)$ is the first-order Bessel function of the first kind. Note that the bracketed term in equation (6.6) represents the received pulse power in watts and that the PPM time slot width in equation (6.6) is related to the channel data rate R by

$$\tau = \frac{\log_2 M}{MR} \qquad (6.8)$$

We now define α_t as a parameter independent of θ_t, P_t, and d_t to be

$$\alpha_t = \eta_t \eta_r \left(\frac{\lambda}{4\pi Z}\right)^2 \left(\frac{\pi d_r}{\lambda}\right)^2 \left(\frac{\eta}{h\nu}\right) L_o M \tau \tag{6.9}$$

Thus, equation (6.6) may be written as

$$K_S(\theta_t, P_t, d_t) = \alpha_t P_t G_t L_t(\theta_t, d_t) \tag{6.10}$$

where $G_t = (\pi d_t/\lambda)^2$. For a given rms pointing error σ_T, equation (6.10) can be expressed as

$$K_S(\theta_t, P_t, d_t) = \left(\frac{\alpha_t P_t}{\sigma_T^2}\right)\left(G_t \sigma_T^2\right)\left(\frac{2J_1\left[\sqrt{G_t \sigma_T^2}\dfrac{\theta_t}{\sigma_T}\right]}{\sqrt{G_t \sigma_T^2}\dfrac{\theta_t}{\sigma_T}}\right)^2 \tag{6.11}$$

In the general M-ary PPM case, the BER is given by equation (3.31). In BPPM, the BER as a function of the received signal counts per slot K_S is given by equation (3.32) as

$$P_E(K_S) = Q\left[\sqrt{\frac{K_S^2}{FK_S + 2FK_B + 2FK_{DB} + 2K_{DS} + 2K_I}}\right] \tag{6.12}$$

Since θ_t is a Rayleigh-distributed random variable, K_S and $P_E(K_S)$ are random variables. In the presence of random pointing error, therefore, the expected value or time-averaged BER is

$$\overline{BER} = \int_0^\infty P_E\left[K_S(\theta_t, P_t, d_t)\right] \frac{\theta_t}{\sigma_T^2} e^{(-\theta_t^2/2\sigma_T^2)} d\theta_t \tag{6.13}$$

Defining the parameters $p = G_t \sigma_T^2$ and $q = \alpha_t P_t / \sigma_T^2$ and letting $u = \theta_t/\sigma_T$ be the normalized instantaneous pointing error, we can express equation (6.13) as

$$\overline{BER} = \int_0^\infty P_E\left[K_S(u, p, q)\right] u e^{(-u^2/2)} du \tag{6.14}$$

where

$$K_S(u, p, q) = pq\left(\frac{2J_1(\sqrt{pu})}{\sqrt{pu}}\right)^2 \tag{6.15}$$

Equation (6.14) demonstrates that the average BER depends only on the parameters p and q. Therefore, if α_t is held constant (i.e., at the same link distance, wavelength, receiver configuration, and other parameters), the same average BER may be maintained for different values of the rms pointing error σ_T by adjusting the transmission power P_t and the transmitter aperture size d_t (hence, G_t) so that p and q remain constant. Thus, there is a tradeoff between transmission power and transmitter aperture diameter. As the rms error increases, transmission power must increase and transmitter aperture size must decrease to keep p and q constant and to maintain the same average link BER for a fixed α_t. If we were to employ equation (6.14) to make a plot of transmission power versus transmitter aperture size for a given average BER and rms pointing jitter, the plot would resemble figure 6.1. For zero pointing jitter, the required power is inversely proportional to the transmitter gain. For nonzero pointing jitter, however, an optimum aperture size exists that minimizes the required transmission power. Using an aperture size larger than the optimum will result in a higher power requirement, because the smaller beamwidth increases the probability of the transmitted optical beam missing the receiving aperture. Suppose that the optimum transmitter aperture size and minimum transmission power d_{opt} and P_{min} can be found for a given link configuration (i.e., constant α_t), average BER, and rms jitter. Then the corresponding parameters p and q are given by

$$p_{opt} = G_{opt}\sigma_T^2 = \left(\frac{\pi d_{opt}}{\lambda}\right)^2 \sigma_T^2 \qquad q_{opt} = \frac{\alpha_t P_{min}}{\sigma_T^2} \tag{6.16}$$

Because these parameters must remain constant for the average BER to remain constant, the optimum transmitter aperture size and minimum transmitter power requirement for any rms

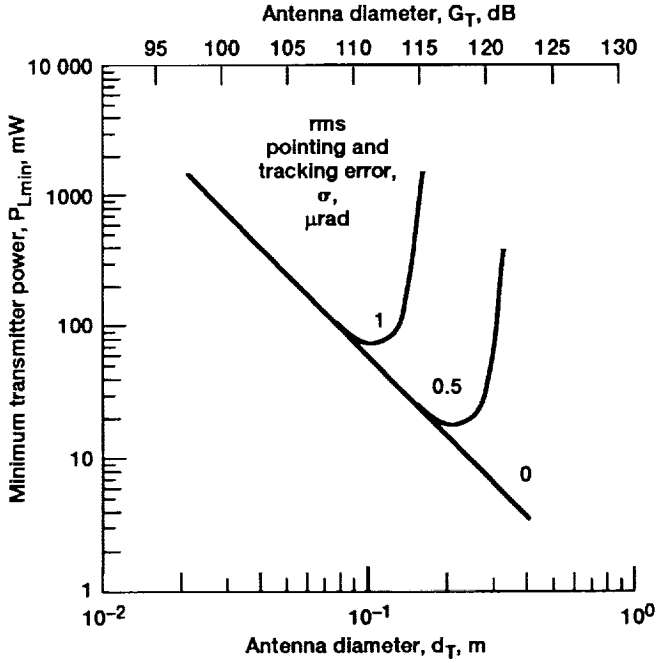


Figure 6.1.—Typical plot of transmitter power and aperture size tradeoff for an optical link in the presence of random pointing error. Direct-detection quaternary pulse position (QPPM); link range, z, 40 000 km; wavelength, λ, 850 nm; receiver aperture diameter, d_r, 30 cm; data rate, R, 200 Mbps; bit error rate, BER, 10^{-9}.

pointing jitter value (assuming α remains constant) can be found from

$$d_{\text{opt}} = \frac{\sqrt{p_{\text{opt}}}}{\sigma_T}\frac{\lambda}{\pi} \qquad P_{\min} = \frac{q_{\text{opt}}}{\alpha_t}\sigma_T^2 \tag{6.17}$$

Hence, the optimum aperture diameter is inversely proportional to the rms pointing error, whereas the minimum power requirement is directly proportional to the square of the rms error (i.e., the pointing error variance). In short, the larger the rms pointing error, the larger the transmission power and the smaller the transmitter telescope size required to maintain the same link BER.

6.2 Pointing and Tracking Error in the Heterodyne-Detection Link

In a heterodyne system, random pointing and tracking errors impose a size constraint on both the transmitter and receiver aperture, since both transmitter pointing and LO and received-signal alignment are affected. The larger the receiver aperture, the more difficult it is to obtain overlapping LO and received signal patterns on the detector surface. If we let θ_t and θ_r serve as the instantaneous pointing error and the received signal and LO misalignment error, respectively, the received signal of photoelectrons per bit can be written similarly to equation (6.6) as

$$\begin{aligned} K_S(\theta_t,\theta_r,P_t,d_t,d_r) &= \left[P_t\eta_t\left(\frac{\pi d_t}{\lambda}\right)^2 L_t(\theta_t,d_t)\left(\frac{\lambda}{4\pi Z}\right)^2 \right. \\ &\quad \left. \times L_r(\theta_r,d_r)\eta_r\left(\frac{\pi d_r}{\lambda}\right)^2 L_o\right]\left(\frac{\eta}{h\nu}\right)T_b \\ &= \beta P_t G_t G_r L_t(\theta_t,d_t)L_r(\theta_r,d_r) \end{aligned} \tag{6.18}$$

where

$$\beta = \eta_t\eta_r\left(\frac{\lambda}{4\pi Z}\right)^2 L_o\left(\frac{\eta}{h\nu}\right)T_b$$

$$G_t = \left(\frac{\pi d_t}{\lambda}\right)^2 \qquad G_r = \left(\frac{\pi d_r}{\lambda}\right)^2$$

The first bracketed term in equation (6.18) represents the average received-signal power, and T_b represents the bit period. The parameter β is independent of θ_t, θ_r, d_t, d_r, and P_t. Again, if we assume that the transmitter aperture is uniformly illuminated and approximate the received signal optical field and the LO optical field by plane waves at the receiver aperture, the pointing-loss factor and LO tracking-loss factor take the form

$$\begin{aligned} L_t(\theta_t,d_t) &= \left[\frac{2J_1(\pi\theta_t d_t/\lambda)}{\pi\theta_t d_t/\lambda}\right]^2 = \left[\frac{2J_1(\sqrt{G_t}\theta_t)}{\sqrt{G_t}\theta_t}\right]^2 \\ &= \left[\frac{2J_1(\sqrt{G_t\sigma_T^2}\theta_t/\sigma_T)}{\sqrt{G_t\sigma_T^2}\theta_t/\sigma_T}\right]^2 \\ L_r(\theta_r,d_r) &= \left[\frac{2J_1(\pi\theta_r d_r/\lambda)}{\pi\theta_r d_r/\lambda}\right]^2 = \left[\frac{2J_1(\sqrt{G_r}\theta_r)}{\sqrt{G_r}\theta_r}\right]^2 \\ &= \left[\frac{2J_1(\sqrt{G_r\sigma_R^2}\theta_r/\sigma_R)}{\sqrt{G_r\sigma_R^2}\theta_r/\sigma_R}\right]^2 \end{aligned} \tag{6.19}$$

where σ_T and σ_R are the rms pointing error and the rms LO tracking (alignment) error, respectively. The received signal counts per bit can be rewritten as

$$\begin{aligned} K_S(\theta_t,\theta_r,P_t,d_t,d_r) &= \left(\frac{\beta P_t}{\sigma_T^2\sigma_R^2}\right)\left(G_t\sigma_T^2\right)\left(G_r\sigma_R^2\right) \\ &\quad \times \left[\frac{2J_1(\sqrt{G_t\sigma_T^2}\theta_t/\sigma_T)}{\sqrt{G_t\sigma_T^2}\theta_t/\sigma_T}\right]^2\left[\frac{2J_1(\sqrt{G_r\sigma_R^2}\theta_r/\sigma_R)}{\sqrt{G_r\sigma_R^2}\theta_r/\sigma_R}\right]^2 \end{aligned} \tag{6.20}$$

Table 4.1 gives the BER's for the various modulation formats under LO shot-noise-limited conditions where the transmitter and LO lasers have narrow line widths and the carrier phase noise is, thus, negligible. For example, the quantum-limited BER for homodyne PSK is given by

$$P_E(K_S) = Q\left[\sqrt{4K_S}\right] \tag{6.21}$$

Since the transmitter pointing error and LO tracking error are assumed to be statistically independent Rayleigh random variables, their joint PDF is simply

$$p(\theta_t,\theta_r) = \frac{\theta_t}{\sigma_T^2}e^{(-\theta_t^2/2\sigma_T^2)}\frac{\theta_r}{\sigma_R^2}e^{(-\theta_r^2/2\sigma_R^2)} \tag{6.22}$$

The expected value of the BER, or the average BER, is then

$$\overline{BER} = \int_0^\infty\int_0^\infty P_E\left[K_S(\theta_t,\theta_r,P_t,d_t,d_r)\right]p(\theta_t,\theta_r)d\theta_t\,d\theta_r \tag{6.23}$$

If we define the parameters from equation (6.20) as

$$p_T = G_t\sigma_T^2 \qquad p_R = G_r\sigma_R^2 \qquad q = \frac{\beta P_t}{\sigma_T^2\sigma_R^2}$$

and change the variables in equation (6.23) by letting $u = \theta_t/\sigma_T$ and $v = \theta_r/\sigma_R$ be the normalized pointing and LO tracking errors, respectively, then the average BER can be written as

$$\overline{BER} = \int_0^\infty \int_0^\infty P_E\left[K_S(u,v,p_T,p_R,q)\right] ue^{-u^2/2}\, ve^{-v^2/2}\, du\, dv \tag{6.24}$$

where

$$K_S(u,v,p_T,p_R,q) = qp_Tp_R\left[\frac{2J_1(\sqrt{p_T}\,u)}{\sqrt{p_T}\,u}\right]^2\left[\frac{2J_1(\sqrt{p_R}\,v)}{\sqrt{p_R}\,v}\right]^2$$

Equation (6.24) demonstrates that the average BER depends only on the parameters p_T, p_R, and q. As in the direct-detection case, optimal transmitter and receiver aperture diameters exist for given β, σ_T, σ_R, and $\overline{BER}$, minimizing the required transmission power. Furthermore, if $\sigma_T = \sigma_R$, the optimal transmitter and receiver aperture diameters are equal to each other. If the optimal aperture sizes and minimum transmission power levels are obtained for a given average BER and rms pointing and LO tracking error, then

$$p_{T\text{opt}} = \left(\frac{\pi d_{t\text{opt}}}{\lambda}\right)^2\sigma_T^2 \quad p_{R\text{opt}} = \left(\frac{\pi d_{r\text{opt}}}{\lambda}\right)^2\sigma_R^2 \quad q_{\text{opt}} = \frac{\beta P_{\min}}{\sigma_T^2\sigma_R^2} \tag{6.25}$$

If we assume that β remains constant, the optimal aperture sizes and minimal transmission power requirement for any rms pointing and LO tracking error can be found from

$$d_{t\text{opt}} = \frac{\sqrt{p_{T\text{opt}}}}{\sigma_T}\frac{\lambda}{\pi} \quad d_{r\text{opt}} = \frac{\sqrt{p_{R\text{opt}}}}{\sigma_R}\frac{\lambda}{\pi} \quad P_{\min} = \left(\frac{q_{\text{opt}}}{\beta}\right)\sigma_T^2\sigma_R^2 \tag{6.26}$$

Note that the minimal power requirement for the heterodyne system is proportional to $\sigma_T^2\sigma_R^2$, whereas the required power for the direct-detection system is proportional only to σ_T^2. A heterodyne system is therefore more sensitive to spatial tracking error than is a direct-detection system operating at the same wavelength. Consequently, the receiver sensitivity advantage offered by heterodyne detection may be offset by the higher pointing losses unless the pointing and tracking error is kept acceptably small.

These equations allow us to determine other BER fluctuation measures. For example, the probability of the instantaneous BER exceeding a specified value (i.e., the burst error probability) or the fraction of time that the BER occurs within specified intervals (i.e., the BER distribution) can be determined by integrating the pointing error PDF using appropriate limits.

6.3 Random Pointing Error Versus Static Pointing and Tracking Error

Of interest is a comparison of the expressions for optimum aperture size in equations (6.17) and (6.26) with those that result if one assumes static, rather than random, pointing and tracking error. In a real environment, of course, the pointing error is random, and the previous expressions yield more accurate results. As mentioned earlier, if the system experiences both static and random pointing errors, the Rician PDF in equation (6.5) must be used in place of the Rayleigh PDF. However, if the rms value of the random error is much smaller than the static pointing error or bias, the analysis can be simplified by assuming that only static error is present.

For the direct-detection system, if we assume that θ_t is constant, the required transmission power for a given signal photocount K_S is, from equation (6.10), equal to

$$P_t = \frac{K_S}{\alpha G_t L_t(\theta_t,d_t)} \tag{6.27}$$

For a given set of link parameters that define α, the transmitter power can be minimized by maximizing the product

$$G_tL_t(\theta_t,d_t) = G_t\left[\frac{2J_1(\sqrt{G_t}\theta_t)}{\sqrt{G_t}\theta_t}\right]^2 = \left[\frac{2J_1(\sqrt{G_t}\theta_t)}{\theta_t}\right]^2 \tag{6.28}$$

Since the Bessel function $J_1(x)$ is maximum at $x = 1.84$, the optimum transmitter aperture gain is

$$\sqrt{G_t}\theta_t = 1.84 \Rightarrow G_{\text{opt}} = \frac{(1.84)^2}{\theta_t^2} = \frac{3.38}{\theta_t^2} \tag{6.29}$$

and the optimum transmitter aperture diameter is

$$d_{\text{opt}} = \sqrt{G_{\text{opt}}}\frac{\lambda}{\pi} = \frac{1.84}{\theta_t}\frac{\lambda}{\pi} \tag{6.30}$$

If equation (6.30) is substituted into $L_t(\theta_t,d_t)$, then the minimum achievable pointing loss equals 0.4, or -4 dB. Thus, the transmission power must be increased 4 dB to maintain the same BER in the presence of static pointing error and the optimal transmitter aperture size. In the case of a fixed pointing error of θ_t—which is equal to the rms value σ_T of a Rayleigh-distributed random pointing error—equation (6.17) may be used to obtain a corresponding optimum aperture diameter ratio of

$$\frac{d_{\text{static}}}{d_{\text{random}}} = \frac{1.84\lambda/(\theta_t\pi)}{\sqrt{p_{\text{opt}}}\lambda/(\sigma_T\pi)} = \frac{1.84}{\sqrt{p_{\text{opt}}}} \tag{6.31}$$

Section 7 shows that the optimum p value for M-ary PPM and a 10^{-6} average BER is $p_{opt} = 0.223$. Therefore, from equation (6.31), the optimum aperture size in the presence of static pointing error is 3.9 times larger than that for random pointing error. This, in turn, yields a static-pointing-error optimum antenna gain that is 15.2 times (11.8 dB) larger than the random-pointing-error optimum antenna gain. Consequently, if the optical-link analysis is conducted under the assumption of static pointing error, rather than under the more realistic random error assumption, the results may differ significantly. Optimistic results are obtained under the static-error assumption because larger apertures are allowed, permitting lower power requirements.

The optimum aperture sizes for the heterodyne system under static pointing and tracking error are found in a manner similar to that used to obtain the optimum aperture sizes of the direct-detection system. Assuming constant θ_t and θ_r and following equation (6.18) yields a required transmission power for a given signal photoelectron count of

$$P_t = \frac{K_S}{\beta G_t G_r L_t(\theta_t,d_t) L_r(\theta_r,d_r)} \tag{6.32}$$

For a given set of link parameters that specify β, the transmission power is minimized by maximizing the product

$$G_t L_t(\theta_t,d_t) G_r L_r(\theta_r,d_r) \tag{6.33}$$

From the previous discussion, this is done by using

$$d_{t\,opt} = \frac{1.84}{\theta_t}\frac{\lambda}{\pi} \quad \text{and} \quad d_{r\,opt} = \frac{1.84}{\theta_r}\frac{\lambda}{\pi} \tag{6.34}$$

The use of these optimum aperture sizes renders a minimum achievable pointing and tracking loss of 0.16 or -8 dB (which is arrived at by substituting equation (6.34) into the product $L_t(\theta_t,d_t)\ L_r(\theta_r,d_r)$). As in the direct-detection case, link analysis conducted under static-error conditions yields optimistic results because larger apertures are allowed than would be with random errors. For example, section 7 demonstrates that $p_{opt} = 0.282$ is the optimum p value for homodyne and heterodyne PSK and a 10^{-6} average BER. According to equations (6.26) and (6.31), the optimum aperture sizes in the presence of static pointing and LO tracking error are about 3.5 times larger than are those under random-error conditions. Consequently, the optimum transmitter and receiver aperture gains in static error are each 12.25 times (10.9 dB) larger than are the random-error optimum aperture gains.

7.0 Comparison of Optical Technologies for High-Data-Rate Return Link From Mars

The manned exploration and settlement of Mars is a major goal of NASA's Space Exploration Initiative (SEI) program. A primary function of a communication system for Mars exploration will be the relay of large amounts of video and science data from the vicinity of Mars back to Earth. Required transmission rates may range from tens to hundreds of megabits per second as human presence grows and exploration activity expands (ref. 15). Because human safety and transportation costs will be principal concerns, there will be a strong need for communication system elements that require low operating power and that are extremely reliable, small, and lightweight.

Although optical systems in general may be able to provide these desirable features, there are a variety of different optical systems to choose from. An optical system suitable for a (geosynchronous-Earth-orbit) GEO-to-GEO crosslink or for an Earth-to-Moon link may not necessarily be the preferred system for a Mars-to-Earth link, because of the much greater link range, higher power requirement, and greater sensitivity to pointing error. Therefore, in this section, several different optical systems are analyzed to assess their feasibility for high-data-rate Mars-to-Earth communications link and to identify key technology performance requirements.

7.1 Optical Implementations and Link Assumptions

The optical link may be implemented using either a direct detection receiver or a heterodyne or a homodyne detection receiver. The direct-detection system has a simpler design and is potentially more reliable than the heterodyne and homodyne systems because neither an LO laser nor optical carrier frequency and phase synchronization is required at the receiver. Since most terrestrial fiber-optic links employ direct-detection, the direct-detection receiver also offers the advantage of a more mature technology base. On the other hand, heterodyne and homodyne systems offer much higher detection sensitivity and background noise immunity than do direct detection systems (i.e., they can achieve a given BER with 5 to 15 dB less power than can the direct-detection system), and these systems can use the same digital modulation and coding schemes as do RF systems. Heterodyne and homodyne systems are more complex than DD systems, however, and their detection sensitivity advantage can be negated by the increased loss they suffer from random pointing and tracking errors and laser phase noise. Of course, the performance of both types of systems also depends on the particular modulation format and the operating wavelength. Because it is unclear which detection method will be preferable for the Mars-to-Earth optical communication link, our discussion will focus on both types of systems. The optical systems analyzed on the following pages include a

(1) 0.532-μm FD (frequency-doubled) Nd:YAG system using direct-detection PPM
(2) 0.532-μm FD Nd:YAG laser system using homodyne BPSK modulation
(3) 0.85-μm GaAs semiconductor laser system using direct-detection PPM
(4) 0.85-μm GaAs semiconductor laser system using heterodyne NCFSK
(5) 1.064-μm Nd:YAG laser system using direct-detection PPM
(6) 1.064-μm Nd:YAG laser system using homodyne BPSK modulation
(7) 10.6-μm CO_2 laser system using homodyne BPSK modulation

Among the various heterodyne and homodyne modulation formats, homodyne PSK was chosen for the Nd:YAG and CO_2 systems since it has the highest detection sensitivity and because Nd:YAG and CO_2 lasers have the requisite narrow linewidth for homodyne operation (i.e., their phase noise losses are small). Because this narrow line width is very difficult to achieve in semiconductor lasers, noncoherent FSK (NCFSK), which has a more relaxed line width requirement, was chosen for the coherent GaAs system. Pulse-position modulation, with M varying between 2 (BPPM) and 256, was assumed for the direct-detection systems. As previously explained, using a higher PPM order for a given data rate permits a reduction in the average power requirement and pulse-repetition frequency (i.e., more bits are transmitted per pulse) and increases peak power and bandwidth requirements (i.e., the pulses are narrower). Direct detection was not considered for the CO_2 laser because of the relatively poor detector quantum efficiency at this wavelength and because of the CO_2 laser's low pulse-repetition rate.

Table 7.1 lists the principal link parameters used in the following analyses. Data rates from 10 to 1000 Mbps were considered. Note that some of the parameters have not been determined because they will be optimized for different values of the rms pointing and tracking error.

All the direct-detection PPM systems are assumed to use a silicon APD and low-noise preamplifier receiver having the parameters shown in table 7.2 and a maximum count-detection strategy. The GaAs heterodyne system and Nd:YAG homodyne systems are assumed to use a silicon PIN photodiode detector, whereas the CO_2 homodyne system uses a radiatively cooled HgCdTe photovoltaic diode. Of course, the homodyne and heterodyne receivers also require the use of appropriate control loops for tracking the phase or frequency of the received optical carrier, adjustment of the LO laser for Doppler shift, and active alignment of received signal and LO beam. All of the systems mentioned also require peripheral hardware for frequency and thermal stabilization of system elements such as transmitter and LO lasers, diode pumping (DP) lasers, FD crystals, and external modulator crystals.

These analyses assume a link between a nominal Mars relay satellite (MRS) and an Earth relay satellite (ERS), with a range of 2.5 AU (astronomical units), or 374 million km (232 million mi), resulting in an associated, one-way propagation delay of 20 min. By comparison, a typical GEO-GEO crosslink has a range of 42 000 km (26 100 mi) (which corresponds to a 60°

TABLE 7.1.—OPTICAL MARS-TO-EARTH LINK SYSTEM PARAMETERS
[Transmitter power and diameter have not been determined.]

Optical link system	Wavelength, μm	Detection	Modulation	Efficiency of transmitter optics, percent	Data rate, Mbps	rms pointing and tracking error, μrad	Link range, AU	Efficiency of receiver optics, percent	Detector quantum efficiency, QE, percent	Other loss[a], dB	Average bit error rate, BER
FD Nd:YAG	0.532	Direct	PPM	80	10 to 1000	0.05 to 0.5	2.5	80	85	3	10^{-6}
FD Nd:YAG	.532	Homodyne	BPSK								
GaAs	.85	Direct	PPM								
GaAs	.85	Heterodyne	NCFSK								
Nd:YAG	1.064	Direct	PPM								
Nd:YAG	1.064	Homodyne	BPSK								
CO_2	10.6	Homodyne	BPSK						50		

[a]Other losses include items such as antenna gain obscuration loss, symbol-synchronization loss, optical-filter loss, carrier phase noise loss for coherent systems, and tracking power loss.
[b]To be determined.

TABLE 7.2.—DIRECT-DETECTION AVALANCHE PHOTODIODE (APD) AND PREAMPLIFIER PARAMETERS

Detector quantum efficiency, η, percent	85
APD and preamplifier capacitance, C, pF	0.10
Preamplifier resistance, R_L	adjusted so that $\tau \approx 2\pi R_L C$
Preamplifier noise equivalent temperature T_{eq}, K	290
APD mean gain, G	200
APD excess noise factor, F	3.183 (5 dB)
APD gain-dependent dark current, I_{DB}, nA	0.10
APD gain-independent dark current, I_{DS}, nA	10.0

separation), resulting in a 140-msec one-way propagation delay. A Moon-to-Earth link has a 385 000-km (239 000-mi) range with a 1.3-sec one-way propagation delay. Since Earth and Mars orbit the Sun, the link range will vary with time, ranging from 0.374 to 2.675 AU.

All the direct-detection systems are assumed to have a 10-m receiver "photon bucket" at the ERS. (The term "photon bucket" refers to a nondiffraction-limited receiving telescope that collects and focuses as much optical energy, or as many photons, as possible onto the detector.) In contrast to the case for the heterodyne and homodyne systems, spatial phase coherence of the beam is not highly significant in direct detection, so aberrations of the receiving optical system are not of great concern as long as the focused spot size does not exceed the photodetector surface area. The FOV of the 10-m DD receiver and its tracking control are designed to keep receiver pointing losses negligible.

The link BER performance requirement is specified as 10^{-6}. In computing the BER, random pointing and tracking errors are modeled as Rayleigh-distributed random variables (i.e., the pointing and tracking error along each of the telescope gimbal axes are assumed to be gaussian distributed and to have zero mean and standard deviation, or rms value σ). Experimental measurements on pointing-control systems developed for optical communications verify that the single-axis pointing error can be accurately modeled by gaussian statistics (ref. 16). As such, the equations in section 6 are applicable, and it is possible to demonstrate that the rms pointing and tracking error σ is a key design parameter of the optical link. As previously discussed, random transmitter pointing error constrains the transmitter aperture size and imposes a power penalty on the direct detection system. In the presence of this random error, an optimum transmitter antenna gain occurs that minimizes the required transmission power necessary to achieve the average BER requirement. In contrast to the situation in the ideal case of zero pointing error, the use of a transmitter aperture larger than the optimum size leads to a higher power requirement rather than to a smaller one. Similarly, in addition to demonstrating random pointing error at the transmitter, heterodyne and homodyne systems (which require precise received-signal beam and LO beam alignment) are also characterized by random tracking error between the received signal beam and the LO beam. Because increases in receiving aperture size lead to increased difficulty in maintaining beam alignment, the heterodyne and homodyne systems' use of a larger receiver aperture does not necessarily lead to a higher SNR. In the presence of random LO tracking error, the receiver aperture size is constrained, and there exists an optimum aperture size that minimizes the required power. In the present analysis, the rms pointing error and rms received signal and LO alignment tracking error are assumed to be equal (i.e., $\sigma_T = \sigma_R = \sigma$). Under this condition, the optimum transmitter and receiver aperture diameters for a particular rms value are equal. Several different values of rms error will be considered, ranging from 0.05 to 0.5 μrad. For a given rms error, optimum link designs (which minimize the required transmission power) will be determined for each of the optical systems.

Finally, additional link margin may be obtained by applying forward error correction (FEC) coding. For DD PPM systems, coding gains on the order of 2 to 3 dB are achievable, depending on the value of M and the decoding approach used (ref. 8). Under LO shot-noise-limited conditions, a coherent system experiences about the same coding gain as does a corresponding RF system. For example, a coherent PSK homodyne system can attain about 8.2 dB of code gain at 10^{-6} BER, using the NASA standard, concatenated, convolutional Reed-Solomon code. However, use of this coding requires that the link be designed to exceed the code threshold for coding gain to be realized (typical code thresholds occur at 10^{-3} channel bit-error-rates). Otherwise, the coding manifests itself only as overhead bits, and link performance is worse than it would have been had no coding been used.

To facilitate our analysis and comparison of the performance capabilities and attendant performance requirements of different optical systems, a review of the principal optical technologies follows.

7.2 Assessment of Key Optical Technologies

Herein, we examine four key optical communication elements: optical transmitters, optical detectors, telescopes, and a tracking and pointing system.

7.2.1 ***Optical transmitters.***—The most critical element of an optical system is the laser transmitter. The general performance requirements for a long-range, space-based laser transmitter are high output power, long lifetime, high efficiency, high reliability, and, for heterodyne and homodyne systems, narrow spectral linewidth. The three principle candidate laser sources are the solid state Nd:YAG (neodymium yttrium aluminum garnet) laser, the semiconductor GaAlAs diode laser (and diode arrays), and the CO_2 laser. Table 7.3 summarizes

TABLE 7.3.—COMPARISON OF LASER TRANSMITTERS

Advantages	Disadvantages
Nd:YAG; wavelength, μm, 1.06 and 0.532	
• High output power: Up to ~15 W at 1.06 μm Up to ~7 W at 0.532 μm 1500 to 2000 W peak power (good for high-order pulse position modulation) • Good reliability with diode array pumping: 40 000 hr, or 5-yr life • Narrow spectral line width: (<1 kHz) suitable for heterodyning and homodyning • High detector QE at 0.532 μm ($70 < \eta < 90$ percent) • Highest antenna gain at 0.532 μm	• Low laser power efficiency: 7 to 10 percent at 1.06 μm with diode pumping, DP 3 to 5 percent at 0.532 μm with DP • Requires external electro-optic, E-0, modulator crystal • Poor detector quantum efficiency, QE, at 1.06 μm ($\eta < 50$ percent • Thermal control required for laser crystal, E-0 crystal, frequency doubled, and DP array
GaAs; wavelength, μm, 0.85	
• High power efficiency: Up to 80 percent for single diode 40 to 50 percent for diode arrays • High reliability: 100 000-hr, 11-yr life expected • High detector QE (80 to 90 percent) • Easily modulated direct current modulation at rates >1 GHz shown • Small size and weight	• Very low output power: 100 mW for single diode 450 mW for diode array 1- to 3-W outpower power expected in 5 to 10 yr • Low peak-to-average-power ratio: (<2 to 4 W) makes them unsuitable for high order PPM • Poor frequency stability and large spectral line width
CO_2; wavelength, μm, 10.6	
• Very high output power: Up to 40 W for RF-excited waveguide lasers • Good power efficiency: (20 to 25 percent) • Suitable for homodyning • Relaxed pointing and tracking • Relaxed optics surface quality	• Relatively short lifetime: 15 000 to 30 000 hr (1.7 to 3.4 yr) • Complex transmitter and receiver external E-0 modulation • Temperature control required for modulator crystal and transmitter and local oscillator lasers • Homodyne detection • Detector cooled to ~100 K • Lowest antenna gain

the important parameters of these three laser types.

7.2.1.1 Diode-pumped Nd:YAG lasers: In the Nd:YAG laser, the light amplification medium is a crystalline rod or slab of YAG, lightly doped with neodymium. Optical energy is used to pump the neodymium ions to a metastable energy state. The Nd:YAG laser operates at a fundamental infrared wavelength of 1.064 μm, and, when used in conjunction with an intracavity FD crystal, can produce second harmonic energy at 0.532 μm (deep green). Both the infrared and harmonic wavelengths fall in the region of greatest photodiode detector sensitivity (i.e., 50 to 85 percent quantum efficiency). In space-based applications, GaAs semiconductor laser-diode arrays provide the optimum technique for pumping the Nd:YAG rod, because the laser-diode emission can be matched to the Nd:YAG absorption bands, resulting in pumping efficiencies (i.e., the ratio of optical output power to electrical pump power) from 7 to 10 percent at 1.064 μm, and from 3 to 5 percent efficiency at 0.532 μm.

The output power capability of a Nd:YAG laser depends on the operating mode (i.e., on whether the mode is continuous wave (CW) or pulsed). Pulse-mode operation is required for direct-detection PPM, whereas heterodyne and homodyne operation require CW operation. Laser pulses are generated by Q switching, cavity dumping, or mode locking. In Q switching, energy builds in the population inversion during a low Q state (high-cavity loss) and then is suddenly removed by a switch to a high Q state (low-cavity loss). TRW has reported peak pulse powers from 1500 to 2000 W with a 10-ns pulse width and from 10- to 20-kHz pulse repetition frequency (PRF) with current laboratory devices. This results in average power levels (peak power × pulse width × PRF) from 150 to 400 mW (TEM_{00} mode output). Commercially available, Q-switched, DP, pulsed Nd:YAG lasers have peak powers of up to 330 W, with 30-nsec pulse duration, 10-kHz PRF, and 100-mW average power (ref. 17). McDonnell Douglas, as a result of research sponsored by the Department of Defense and NASA, recently reported the generation of 3750-W peak power with 200-μsec pulse duration and 35-W average power from a DP Nd:YAG slab laser (ref. 18). The PRF in this case, however, works out to only 47 Hz—much too slow for

high-rate PPM. A space-qualified, DP Nd:YAG laser has been developed by McDonnell Douglas for use in a Department of Defense space crosslink subsystem.

In the cavity-dumped mode, energy is first built up in the oscillating photon field in a high-Q cavity (low loss) and is then "dumped" by a brief deflection of the beam out of the laser cavity. Peak powers of 50 W with a 10-nsec pulse width and a 300-kHz PRF (150-mW average power) have been reported by TRW. Commercial devices demonstrated at the 1989 Conference on Lasers and Electro-Optics (CLEO) had average power levels of 4 W at 1.06 μm and of 2 W at 0.532 μm.

In mode-locking, the pulse width is inversely related to the range of frequencies over which the laser can be made to oscillate in phase simultaneously. This mode can produce the highest PRF's (gigahertz range). Recently, Coherent Laser Groups reported the production of 6 W of average power at 0.532 μm by frequency doubling the output from a mode-locked Nd:YAG laser operating at 1.064 μm, with 24-W average power, 100-ps pulsewidth, and 76-MHz PRF (ref. 19). The FD crystal was lithium triborate (LBO).

For homodyne and heterodyne operation, the Nd:YAG laser offers a narrow line width and good spectral purity over semiconductor lasers. Output power for commercial CW DP Nd:YAG lasers ranges up to about 80-mW TEM_{00} for 0.532 μm and up to 1-to 2-W TEM_{00} for 1.064 μm (ref. 17). Such lasers may be capable of attaining power levels of 10-W CW within the next 5 yr. Line widths of less than 1 kHz can currently be achieved. The chief disadvantage of using the Nd:YAG laser in heterodyne systems is that an external electro-optic or acousto-optic modulator is needed to modulate the laser; such modulators can introduce additional signal power loss and carry a large prime-power requirement ($\approx$ 100 W). The lifetime of the DP Nd:YAG laser is determined by the life of the semiconductor pumping diode array, which is estimated to be about 40 000 hr.

7.2.1.2 Semiconductor lasers: Semiconductor lasers offer the following advantages:

(1) High efficiency (up to 80 percent for a single diode)
(2) High reliability (extrapolated 100 000-hr lifetime)
(3) High detection sensitivity (0.8-μm wavelength)
(4) High modulation bandwidth (direct-current modulation in excess of 1 GHz)
(5) Ruggedness and small size.

The major disadvantage of semiconductor lasers is their small output power. Commercial GaAs laser diodes suitable for direct detector systems have an average output power in the 75- to 100-mW range (TEM_{00} mode). Although these devices emit only in a single transverse mode, they usually oscillate on many longitudinal modes, and thus are characterized by poor frequency stability and large phase fluctuations at their output.

Although these large line width lasers are unsuitable for heterodyne and homodyne systems, feedback and coupled-cavity techniques can be employed to improve their frequency stability and to reduce their laser line width. However, the price of such improvement is greater complexity and smaller output-power capability. The Massachusetts Institute of Technology Lincoln Laboratory (MIT-LL) recently reported the development of the first space-qualified, coherent semiconductor laser transmitter (ref. 20) that contains four 30-mW GaAlAs diode lasers that operate at 0.86 μm. The transmitter serves as a key subsystem for the laboratory's coherent Laser Intersatellite Transmission Experiment (LITE) program, which can deliver 220 Mbps over a 37 000-km (23 000-mi) crosslink.

The limited output power of single-laser diodes has spawned research directed at developing techniques for combining the output power of multiple laser diodes into a single near-diffraction-limited output beam. The power-combining techniques so far developed here have fallen into two categories: (1) noncoherent combining techniques, which combine emissions of different wavelengths and/or polarizations without maintaining phase coherence among the sources, and (2) coherent combining techniques, which maintain phase coherence among the separate emissions. In the area of monolithic, linear laser-diode phased arrays and surface emitting arrays, TRW recently developed a noncoherent diode array with 450-mW CW single (transverse) mode power, and the company expects to develop a coherent device with 1-W mode power in the near future. Currently, most laser diodes have power outputs of far less than 100 mW. In general, the development of long-life, coherent GaAs devices whose output power exceeds 1 W and which can produce a high-quality, single-lobed far-field beam remains a significant technological challenge.

7.2.1.3 Carbon dioxide lasers: The CO_2 gas laser, which operates at a 10.6-μm nominal wavelength in the far infrared field, was an early candidate for use in space communications. NASA began work on CO_2 lasercom systems in the late 1960's and on a planned experiment for the ATS-F and ATS-G technology satellites. However, the experiment was subsequently cancelled as a result of funding problems (ref. 21). NASA's CO_2 system development continued into the early 1970's. Over the past several years the European Space Agency (ESA) and the West Germans have continued to work on CO_2 laser-system development (refs. 22 and 23).

The major advantage of the CO_2 laser is its high output power—it can generate the highest CW power of any laser. It also has relatively high efficiency (up to 20 percent). The laser can operate in a pulsed mode, but not at high enough repetition rates to be useful for communications. There are several different configurations possible with this laser. Continuous wave power can range from a few watts for the sealed-tube and waveguide configurations to tens of kilowatts for the high-power, transverse gas flow industrial lasers used for cutting and drilling. The laser's active medium is a gas mixture of carbon dioxide, nitrogen, and helium. The low-energy, long-wavelength emission results from transitions between different vibrational and rotational modes of the CO_2 molecule.

In the past, the limited lifetime of conventional CO_2 lasers, which are electrical gas-discharge devices, has been their major drawback for use in space. This short lifetime results from both the relatively brief lifespan of vacuum seals, cathodes,

and anodes, and the decomposition and eventual exhaustion of the CO_2 (and, hence, of the power output) that results from the electrical-discharge-gas interaction in a sealed-gas-tube system.

With the development of compact RF-excited CO_2 waveguide lasers, however, lifetimes may be extended to those required for space application. The life expectancies of selected laboratory devices range from 30 000 to 50 000 hr, (or from 3 to 6 yr) (ref. 22). These lasers use an external RF source, rather than an electrical discharge, to excite the gas. They also use a small tube of a few millimeters in diameter to contain the gas mixture; the small tube improves the lasing efficiency of these lasers and allows them to acquire high gain despite their short length. For example, a typical commercial waveguide laser may be only 40-cm long and yet can output 8 W of power (ref. 17). Such lasers may also use a high-pressure gas reservoir to periodically replenish the gas in the tube and further extend their life. The power outputs of commercially available waveguide lasers range from under 1 W to 25 W (single mode) (ref. 17).

External electro-optic modulation techniques have been developed for high-data-rate modulation of the CO_2 laser. For example, testing of an ESA laboratory model of a CO_2 homodyne PSK transceiver indicates that phase-modulated bandwidths of up to 5 GHz can be obtained with a cadmium telluride (CdTe) electro-optic modulator. In the ESA test, the modulator (on the average) produced 1.3 W of modulated output power, with 6.2 W of optical input power (ref. 22).

Another characteristic of CO_2 lasers that has hindered their application in space is the relatively small aperture gain that they experience as a result of their long wavelength. For the 2.5-AU Mars-to-Earth optical link, however, the lasers' long wavelengths may prove advantageous, because their wider beamwidths result in less stringent pointing and tracking requirements than for other optical systems. In addition, good "optical quality" surfaces require surface tolerances of 1 μm, which, for the CO_2 laser, can be easily achieved and measured with conventional machining technology. CO_2 laser "optics" are generally made of a metal, such as selenium or germanium, since conventional silica glass is opaque at 10.6 μm. Their long wavelengths may also make CO_2 systems amenable to large-diameter segmented mirrors that can be assembled in space.

7.2.2 ***Optical detectors.***—The optical detector's general-performance requirements are high quantum efficiency, high bandwidth, high gain, and low noise. Optical detectors commonly used in communication include the vacuum photomultiplier tube (PMT), p-type, intrinsic n-type (PIN) photodiode, and avalanche photodiode (APD).

A PMT consists of a photocathode, a series of electrodes (called dynodes), and a collecting anode. The dynodes are assigned progressively higher potentials with respect to the cathode; a typical potential difference between adjacent dynodes is 100 V. Electrons, emitted from the photocathode by incident radiation, are accelerated toward the first dynode by the applied field. The impact of these electrons on the dynode creates secondary electrons, which are then accelerated toward the next dynode. This process is repeated at each dynode until the initial photocathode current is amplified by a very large factor. Typical PMT's have gains from 10^5 to 10^6. They also have very low gain noise (i.e., each electron emitted from the photocathode generates approximately the same number of secondary electrons). Major disadvantages, however, include their poor quantum efficiency (from 1 to 20 percent) at the high visible and near-infrared wavelengths (on which the Nd:YAG and GaAs lasers transmit) and their low bandwidth (or slow response time), which results from the time consumed by the electrons as they traverse the dynode chain. Among the PMT's other disadvantages are bulkiness, high-voltage requirements (from 1000 to 3000 V), and a need for extensive thermal cooling.

Semiconductor photodetectors (PIN diodes and APD's) offer the advantages of high quantum efficiency (from 60 to 90 percent), high bandwidth (>1 GHz), high reliability, and small size. PIN diodes, which possess unity internal gain, are applicable in heterodyne and homodyne systems in which a strong LO laser provides the effective gain necessary to achieve near quantum-limited performance (i.e., where LO shot noise dominates the detector thermal noise). Current silicon PIN diodes have quantum efficiencies of 60 to 90 percent at wavelengths of 0.532 and 0.85 μm and efficiencies of 40 percent at the Nd:YAG fundamental wavelength of 1.064 μm. An AlGaAs PIN device, custom made by MIT-LL, has an 85 percent quantum efficiency at 0.85 μm (ref. 24).

In direct-detection systems, thermal noise is overcome by using an APD with large internal gain. Unfortunately, because the impact ionization process in an APD is random, the multiplication gain is random, and thus avalanche gain noise degrades the detector SNR from the shot-noise-limited value. APD gain noise is characterized by its excess noise factor F (ratio of the mean-squared gain to the mean gain squared). An ideal APD with constant gain would have $F = 1$. Currently, reach-through Si APD's have gains of 100 to 300, quantum efficiencies of 80 to 90 percent at 0.532, wavelengths of 0.85 μm (40 to 50 percent at 1.064 μm), and excess noise factors of 2.8 to 3.2 (ref. 24). As APD's with lower excess noise factors are developed, the sensitivity of direct-detection systems will begin to approach that of heterodyne-detection systems. Typical APD bandwidths range from 100 MHz to 4 GHz at visible and near-infrared wavelengths. Separate-absorption-and-multiplication-region APD's (SAM- APD) and multiple-quantum-well APD's (MQW-APD) are being developed for wide-bandwidth detection at longer wavelengths.

7.2.3 ***Telescopes.***—The expensive low-loss lens materials and comparatively greater weights associated with refractive optics have rendered reflective optics that have a Cassegranian or Newtonian configuration the preferred choice for large telescopes (>10 cm). To reduce weight, TRW is developing telescope mirrors with metal foam cores; foams can be reduced to a density equal to less than 2 percent that of a solid with

an equivalent thickness. For example, the University of Arizona has incorporated metal foam-core mirrors in a 30-cm-aperture Cassegranian telescope that weighs only 4.5 kg. Lightweight optics are important for the Mars-to-Earth optical link because the 20-min, one-way propagation delay will require separate transmitter and receiver telescopes and gimbals for both the MRS and ERS. The MRS will need to have its receiver optics pointed such that it can both receive signals transmitted from the ERS 20 min before, and transmit signals to the position the ERS will have attained 20 min into the future. Technology being developed for large space-borne telescopes, such as the Hubble Space Telescope and the large deployable reflector (LDR), will contribute to the development of apertures of similar size for Mars communications. For example, the Hubble 2.4-m (7.9-ft) primary mirror features a honeycombed glass internal structure, which has made possible the mirror's comparatively light weight of 818.2 kg (1800 lb). The 20-m (65.6-ft) aperture LDR, to be deployed early in the 21st century, will be assembled in space and will consist of at least 60 mirror segments. In addition to overcoming weight and cost problems, large-diameter optical telescopes must also possess high-surface quality, good thermal stabilization, high rigidity, and anticontamination shielding to be feasible in space applications.

7.2.4 ***Tracking and pointing system.***—The required RMS pointing error is the critical parameter that drives the tracking and pointing control system design. The tracking and pointing subsystem must be capable of submicroradian accuracy for high-data-rate (> 10 Mbps) operation as the following analysis will show. Furthermore, the large propagation delay of a Mars-to-Earth optical link requires that the acquisition and tracking operation be accomplished in an open-loop mode rather than in a closed-loop mode. Because platform disturbance is usually the dominant source of tracking error, careful consideration must be given to spatial tracking loop design and methods of suppressing satellite vibration. High-tracking loop bandwidth (a few kilohertz) is necessary to attenuate the high-frequency components in the platform-disturbance spectrum. Measurements of LANDSAT platform jitter indicate that rms pointing errors of about 0.5 μrad are currently typical (ref. 24). Ball Aerospace has developed and tested an experimental tracking and pointing subsystem capable of 0.1-μrad pointing accuracy in the presence of simulated LANDSAT platform disturbances (ref. 16). Measurements of the pointing distribution of this experimental subsystem also verified that single-axis pointing errors along the azimuthal and elevation telescope gimbal axes can be modeled accurately by gaussian statistics. As another point of comparison, the sophisticated pointing system aboard the 12.5-ton Hubble Space Telescope is capable of holding an aperture of a scientific instrument to an accuracy of 0.034-μrad (0.007-arc sec) rms for the duration of an astronomical observation (ref. 25). The use of this complex system, however, could significantly increase the weight and cost of a relay satellite.

Adaptive platform-jitter compensation may offer a more feasible means of attaining extremely precise pointing accuracy in such satellites. NASA's Instrument Pointing System (IPS) uses this technique. The IPS has a 2-m-diameter equipment platform and is designed to aim a 440- to 14 000-lb payload to an accuracy of about 5 μrad while based on an unstable spacecraft such as a shuttle or Space Station Freedom. To reject the effect of base disturbances on the platform, IPS employs linear actuators driven by computers that receive attitudinal and vibrational inputs from optical and inertial sensors. For example, a star sensor and gyroscope continuously measure platform attitude while accelerometers measure rapid changes in attitude. The computers then process these inputs and provide instructions to the actuators for virtually instantaneous correction of telescope aim in compensation for undesired base perturbations. Of course, the principal disadvantage of such a control system is its complexity: use of the IPS requires the implementation of 200 000 software instructions, compared with the 6000 instructions typical of a satellite not equipped with the system.

Tracking and pointing subsystem performance, expressed in terms of the rms pointing error σ, is perhaps the most important consideration in the design of a Mars-to-Earth optical link. As will be shown herein, the rms error constrains the telescope aperture size and, therefore, affects the transmission power requirement (i.e., the selection of the laser source) and the modulation-detection scheme.

7.3 Analyses of Optical Links

From section 6, the average BER for a direct-detection system, assuming the pointing and tracking errors are Rayleigh distributed, is

$$\overline{BER} = \int_0^\infty P_E(u,p,q)ue^{(-u^2/2)}du \tag{7.1}$$

Similarly, the average BER for heterodyne and homodyne systems is

$$\overline{BER} = \int_0^\infty\int_0^\infty P_E(u,v,p_T,p_R,q)ue^{(-u^2/2)}ve^{(-v^2/2)}du\ dv \tag{7.2}$$

The form of $P_E()$ in these equations will depend on the specific modulation format. In any case, the value of equation (7.1) depends only on the parameters p and q and the value of equation (7.2) depends only on the parameters p_T, p_R, and q. In general, for a given average BER, there exists an infinite number of solutions $\{p,q\}$ that satisfy equation (7.1); likewise, there exists an infinite number of solutions $\{p_T, p_R,q\}$ that satisfy equation (7.2). By using a computer to generate solutions to equation (7.1) and examining these solutions, we find a p value, p_{opt}, and an associated minimum q value, q_{opt}. Likewise, by examining the solutions to equation (7.2), we find a minimum q value, q_{opt}, when $p_T = p_R = p_{opt}$. (In other words, the minimum value for q occurs when p_T and p_R are equal.) This is significant because, if we follow section 6 and let $\sigma = \sigma_T$, it is apparent that the parameters p and q for the direct-detection system are related

to the actual link parameters by

$$p = G_t\,\sigma^2 = \left(\frac{\pi d_t}{\lambda}\right)^2 \sigma^2 \qquad q = \frac{\alpha P_t}{\sigma^2} \tag{7.3}$$

with

$$\alpha = \eta_t \eta_r \left(\frac{\lambda}{4\pi Z}\right)^2 \left(\frac{\pi d_r}{\lambda}\right)^2 \left(\frac{\eta}{h\nu}\right) L_o \frac{\log_2 M}{R} \tag{7.4}$$

Also from section 6 and letting $\sigma = \sigma_T = \sigma_R$, the parameters p_T, p_R, and q for the heterodyne and homodyne systems are related to the link parameters by

$$p_T = \left(\frac{\pi d_t}{\lambda}\right)^2 \sigma^2 \qquad p_R = \left(\frac{\pi d_r}{\lambda}\right)^2 \sigma^2 \qquad q = \frac{\beta P_t}{\sigma^4} \tag{7.5}$$

with

$$\beta = \eta_t \eta_r \left(\frac{\lambda}{4\pi Z}\right)^2 L_o \left(\frac{\eta}{h\nu}\right) \frac{1}{R} \tag{7.6}$$

Note from the expression for q in equations (7.3) and (7.5) that, for given values of σ and α (or β), the minimum q value, q_{opt}, corresponds to the minimum transmission power, $P_t = P_{min}$. Note also, that the associated p value, p_{opt}, corresponds to the transmitter aperture diameter that yields this minimum transmission power level. (For the heterodyne and homodyne systems, p_{opt} corresponds to both transmitter and receiver aperture sizes, which are equal since $\sigma_T = \sigma_R$ is assumed.) Consequently, for any rms error σ and set of link parameters that specify α (or β), the optimal aperture size or sizes d_{opt} that yield the optimal (minimum) transmission power P_{min} can be found directly from the optimum solution pair (p_{opt},q_{opt}) via equations (7.3) or (7.5). Table 7.4 lists the optimum (p,q) parameters for M-ary PPM DD and for each of the binary heterodyne and homodyne modulation schemes. From equations (7.1) and (7.2), it is apparent that these values depend only on the particular modulation type and the average BER value; they were found by substituting the appropriate, instantaneous BER expression for $P_E(\ )$ in equations (7.1) and (7.2) and by generating solutions for $\overline{BER} = 10^{-6}$. This analysis analysis is concerned with the entries for direct detection PPM, homodyne PSK, and heterodyne NCFSK. Note from the expressions for p in equations (7.3) and (7.5) that the optimum aperture size necessary to minimize transmission power for a particular system is dictated by the rms error and the wavelength. The parameter p can be expressed in terms of the rms error-to-beamwidth ratio by rewriting p in equation (7.3) or equation (7.5) as

TABLE 7.4.—OPTIMUM (p,q) VALUES OF OPTICAL DETECTION SCHEMES FOR RAYLEIGH-DISTRIBUTED POINTING AND TRACKING ERROR AND 10^{-6} BIT ERROR RATE

Detection	Modulation	Optimum pointing parameters	
		p_{opt}	q_{opt}
DD	$M = 2$	0.223	518.8
DD	$M = 4$		539.8
DD	$M = 8$		542.0
DD	$M = 16$		539.2
DD	$M = 32$		537.6
DD	$M = 64$		536.6
DD	$M = 128$		538.5
DD	$M = 256$		541.8
Homodyne	PSK	.282	176.3
Heterodyne	PSK	.282	352.7
Differential	PSK	.273	463.7
Heterodyne coherent	FSK	.282	705.3
Heterodyne	NCFSK	.267	927.1
Heterodyne	ASK	.282	1411.0

$$p = \pi^2 \left(\frac{\sigma}{\lambda/d}\right)^2 \tag{7.7}$$

The optimum p values shown in table 7.4 correspond to rms error-to-beamwidth ratios of 0.15 to 0.17. In other words, the transmission power is minimized when σ can be held to 15 to 17 percent of the transmitted beamwidth. The optimum aperture size for the different optical systems at given values of σ can be calculated using equation (7.7) and the p values in table 7.4. These aperture sizes are shown in table 7.5. Values listed for the direct-detection systems refer to the transmitter aperture diameter only, whereas values for the heterodyne and homodyne systems refer both to the transmitter and receiver apertures. Note that, for a given wavelength, the optimum aperture size is inversely proportional to σ; doubling σ halves the aperture size. Using an aperture size other than the optimum size indicated in table 7.5 results in a higher transmission power requirement for the same BER performance. A smaller aperture size increases the power requirement because the smaller transmitted beamwidth reduces the probability of the beam illuminating the receiver. For example, figures 7.1 to 7.3 illustrate the transmission power and aperture size tradeoff for three 100-Mbps Mars-to-Earth optical-link implementations for different values of σ. These plots were constructed using equations (7.1) to (7.6), the optimum q values in table 7.4, and the parameters in tables 7.1 and 7.2. These plots demonstrate the existence of an optimum aperture size for a particular σ.

Besides constraining the aperture size, the random pointing and tracking error also imposes a power penalty or pointing and tracking loss on the system. In other words, for a given aperture size, more power will be required to achieve a given BER in the presence of random pointing and tracking error than will be necessary under zero-error conditions. The power

TABLE 7.5—OPTIMAL APERTURE SIZE VERSUS rms POINTING AND TRACKING ERROR FOR OPTICAL IMPLEMENTATIONS

Optical implementation	Optimal aperture size, cm					
	rms pointing and tracking error,[a] σ, μrad					
	0.05	0.1	0.2	0.3	0.4	0.5
0.532-μm FD Nd:YAG BPPM[b]	160	80	40	27	20	16
0.532-μm FD Nd:YAG Homodyne PSK[c]	180	90	45	30	23	18
0.85-μm GaAs BPPM[b]	256	128	64	43	32	26
0.85-μm GaAs Heterodyne NCFSK[c]	280	140	70	47	35	28
1.064-μm Nd:YAG BPPM[b]	320	160	80	53	40	32
1.064-μm Nd:YAG Homodyne PSK[c]	360	180	90	60	45	36
10.6-μm CO_2 Homodyne PSK[b]	3600	1800	900	600	448	360

[a]Pointing error and local oscillator tracking error assumed equal for heterodyne and homodyne systems (i.e., $\sigma = \sigma_t = \sigma_R$).
[b]Value denotes optimal diameter of transmitter aperture.
[c]Value denotes optimal diameter of both transmitter and receiver apertures.

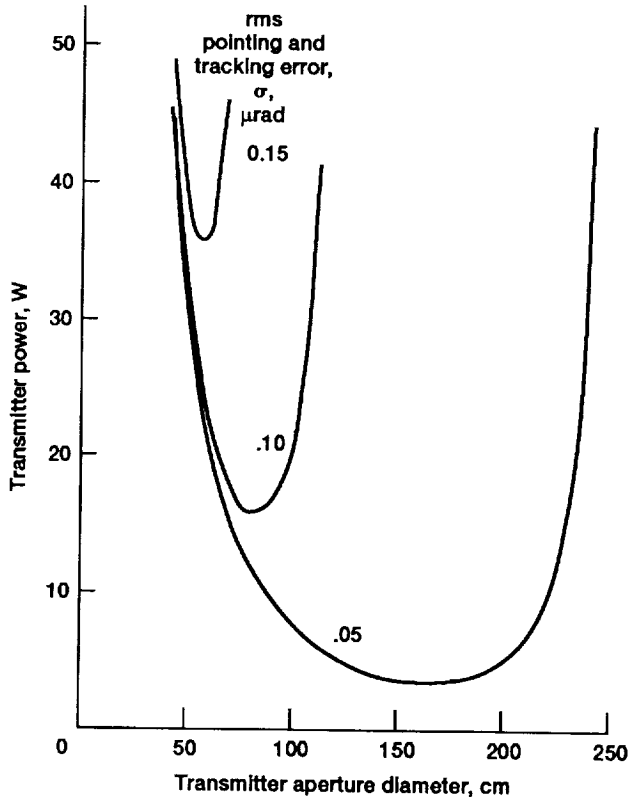


Figure 7.1.—Transmitter power and aperture size tradeoff for Mars binary pulse-position-modulation (BPPM) optical link. Bit error rate, BER, 10^{-6}; wavelength, λ, 0.532 μm; data rate, R, 100 Mbps; receiver aperture diameter, d_r, 10 M; link range, z, 2.5 AU; quantum efficiency, QE, 0.85.

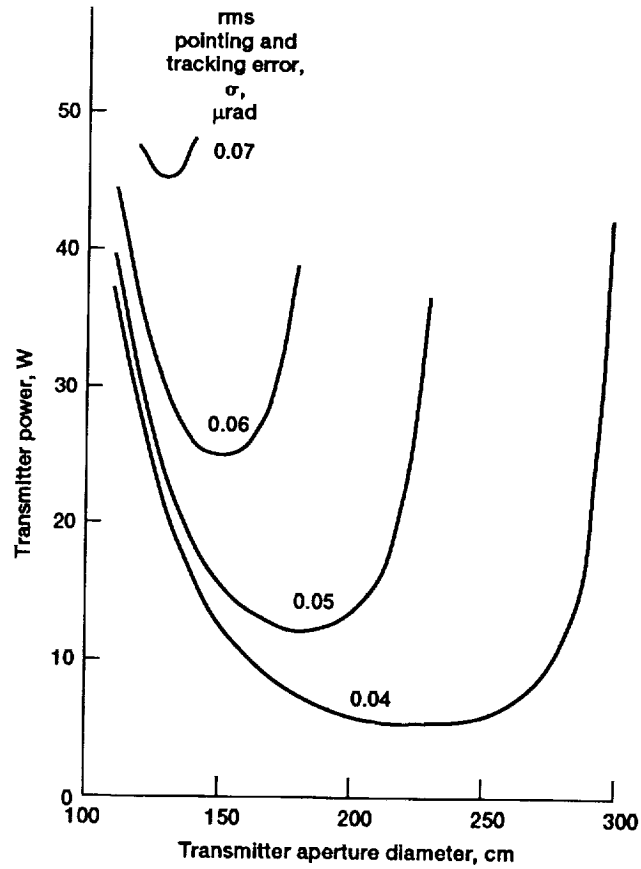


Figure 7.2.—Transmitter power and aperture size tradeoff for Mars homodyne PSK link. Bit error rate, BER, 10^{-6}; wavelength, λ, 0.532 μm; R, 100 Mbps; transmitter aperture diameter, d_t, equals receiver aperture diameter, d_r; link range, Z, 2.5 AU; quantum efficiency, QE, 0.85.

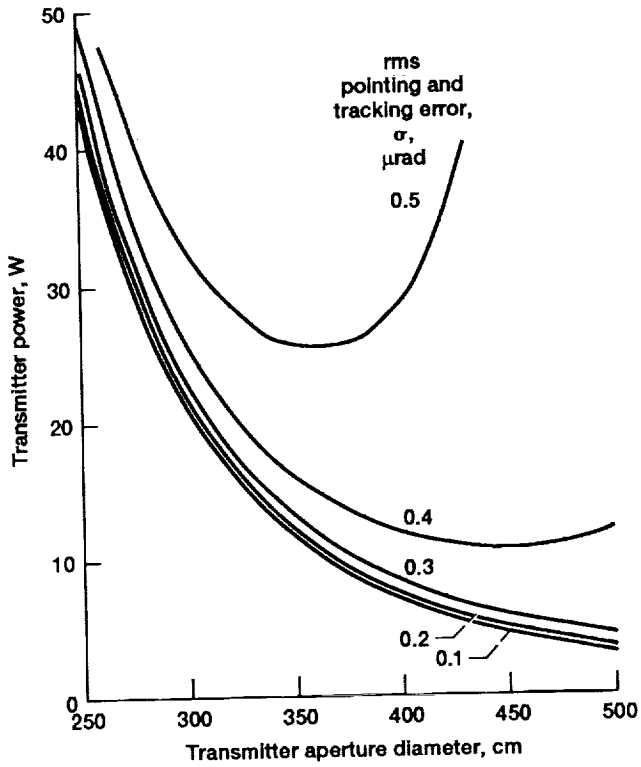


Figure 7.3.—Transmitter power and aperture size tradeoff for Mars homodyne CO_2 link. Bit error rate, BER, 10^{-6}; wavelength, λ, 10.6 μm; R, 100 Mbps; transmitter aperture diameter, d_t, equals receiver aperture diameter, d_r; link range, z, 2.5 AU; quantum efficiency, QE, 0.50.

penalty or pointing and tracking loss can be defined as the ratio of the required signal power necessary to achieve a given BER in the presence of random pointing and tracking error, to that necessary to achieve the same BER under ideal, zero pointing-error conditions using the same aperture sizes.

For the direct detection system, the power penalty can be found as follows. From equation (7.3), the required transmission power with random error present can be expressed as

$$P_t = \frac{q}{\alpha}\sigma^2 = \frac{q}{\alpha}\left(\frac{p}{G_t}\right) = \frac{pq}{\alpha G_t} \tag{7.8}$$

From equation (6.10), it is apparent that the required power with zero pointing error is

$$P_{\text{ideal}} = \frac{K_S}{\alpha G_t} \tag{7.9}$$

The power penalty is therefore

$$L_p = \frac{P_t}{P_{\text{ideal}}} = \frac{pq}{K_S} \tag{7.10}$$

Note that, with an optical link of optimized power budget (i.e., an optical link using minimum transmission power), the product $pq = p_{\text{opt}}q_{\text{opt}}$ is a constant for a given detection-modulation type and average BER). For example, if we use the values for BPPM in table 7.4, $p_{\text{opt}}q_{\text{opt}} = (0.223)(518.8) = 115.7$. Also, if we use the APD parameters from table 7.2 in equation (3.32), the required signal count per bit for 10^{-6} BER is $K_S = 80$ (a time slot width of 5 nsec corresponding to $R = 100$ Mbps is assumed in computing equation (3.32)). From equation (7.10), it is apparent that the power penalty is therefore $L_p = 115.7/80 = 1.446$ (1.6 dB). In other words, if the optimum size transmitter aperture depicted in table 7.5 is used, the power penalty for $R = 100$ Mbps is equal to 1.6 dB for all of the direct-detection BPPM systems. Note that the power penalty with an optical link of optimized power budget is independent of the wavelength and rms error (the optimum aperture size does, however, depend on both these parameters).

The power penalties for the homodyne PSK and NCFSK systems are found in a similar manner. From equation (7.5), it is apparent that the required transmitter power with random pointing and tracking error is

$$P_t = \frac{q}{\beta}\sigma^4 = \frac{q}{\beta}\frac{p_T}{G_t}\frac{p_R}{G_r} \tag{7.11}$$

If we assume equal transmitter and receiver aperture sizes, from equation (6.18), it is apparent that the required power with no pointing and tracking error is

$$P_{\text{ideal}} = \frac{K_S}{\beta G_t G_r} \tag{7.12}$$

The power penalty is therefore

$$L_P = \frac{P_t}{P_{\text{ideal}}} = \frac{qp_Tp_R}{K_S} \tag{7.13}$$

In the optimized optical-link power budget, $p_T = p_R = p_{\text{opt}}$, $q = q_{\text{opt}}$, and the power penalty is

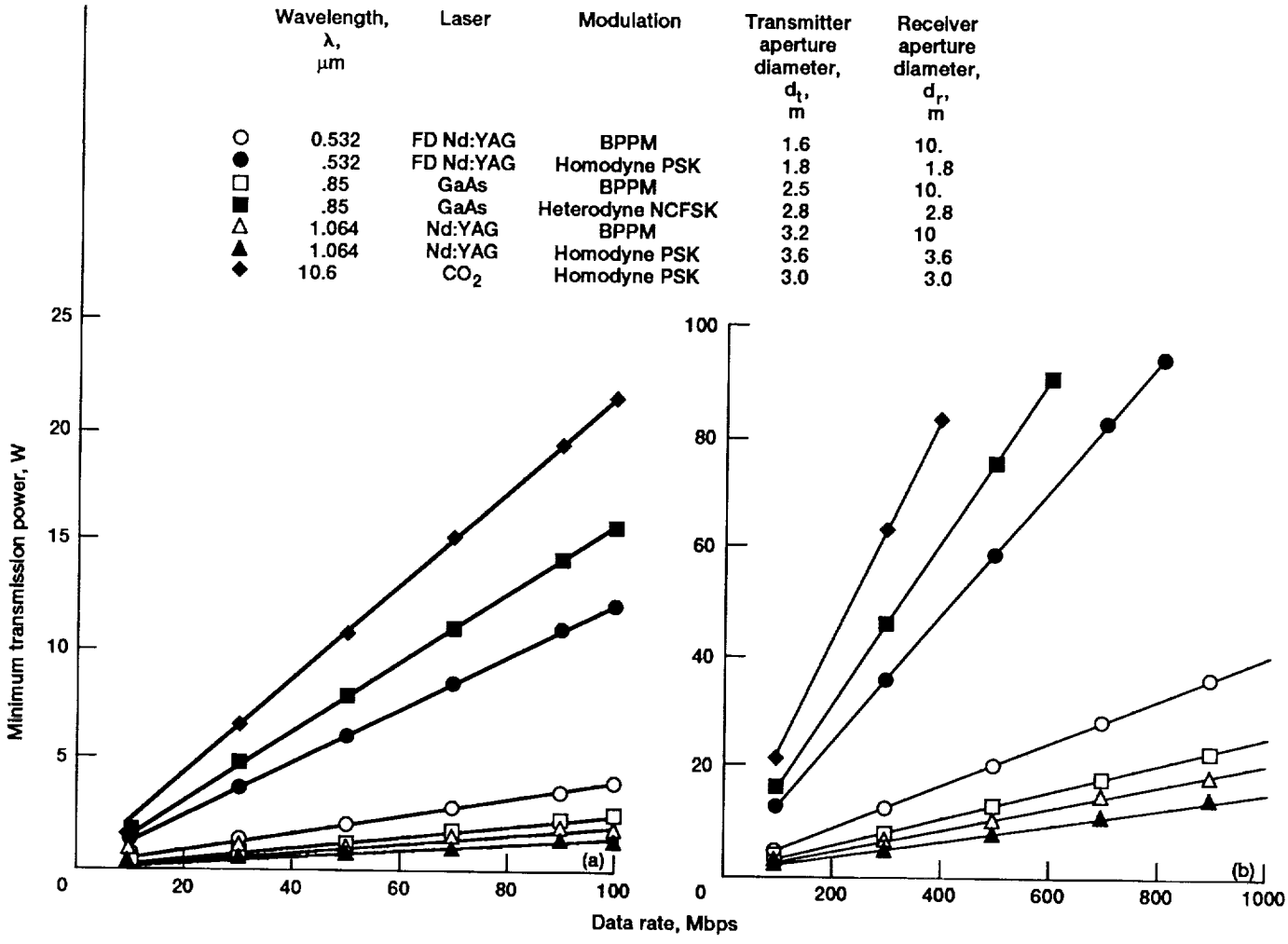
$$L_p = \frac{p_{\text{opt}}^2 q_{\text{opt}}}{K_S} \tag{7.14}$$

The power penalty under optimized conditions, therefore, depends only on the type of heterodyne modulation and the BER assumed for equation (7.2). If we assume a 10^{-6} BER, the required signal counts per bit for homodyne PSK and NCFSK are $K_S = 6$ and $K_S = 27$, respectively (table 4.1). For the (p,q) values from table 7.4, the power penalty from equation (7.14) is 3.7 dB for the homodyne PSK systems and 3.9 dB for the GaAs NCFSK system. Note that these values

represent the minimum power losses that can be achieved for a particular wavelength and rms error. As illustrated in figures 7.1 and 7.2, pointing losses increase rapidly for aperture sizes larger than the optimum value.

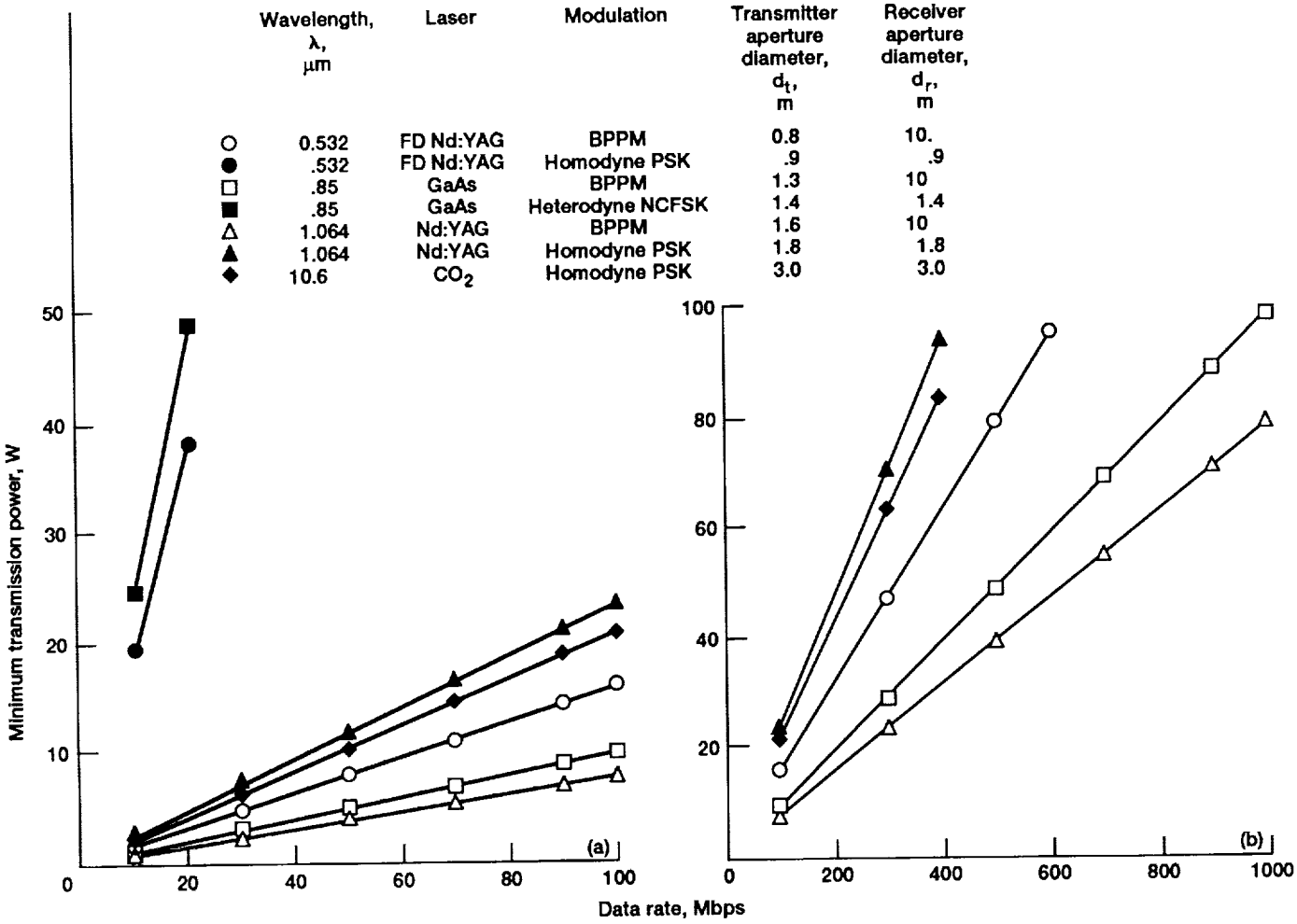
8.0 Summary of Results

The optimum q values in table 7.4 and the link parameters in table 7.1 were used to compute α and β; then minimum transmission power requirements were computed as a function of the link data rate and σ for the various optical systems. These power requirements are shown in figures 7.4 to 7.9. The (a) plots are for data rates in the 10- to 100-Mbps range, whereas the (b) plots are for the 100- to 1000-Mbps range. Some of the graphs do not show the curves for all seven optical systems because the curves lie outside the limits of the graph for some systems. The CO_2 system curves were computed using an aperture diameter of 3.0 m for the transmitter and receiver, instead of the optimum sizes listed in table 7.5 (i.e., the CO_2 curves constitute a suboptimum design). Although the CO_2 optimum aperture sizes result in extremely low powers (several milliwatt) for the small rms error, these aperture sizes are so large that they would be impractical in a real system. Therefore, the smaller 3.0-m aperture was chosen for this analysis at the expense of a higher power requirement. Figures 7.4 to 7.9 illustrate the high sensitivity of the various optical systems to the rms pointing and tracking error. The rms error is therefore a critical parameter in defining the system.



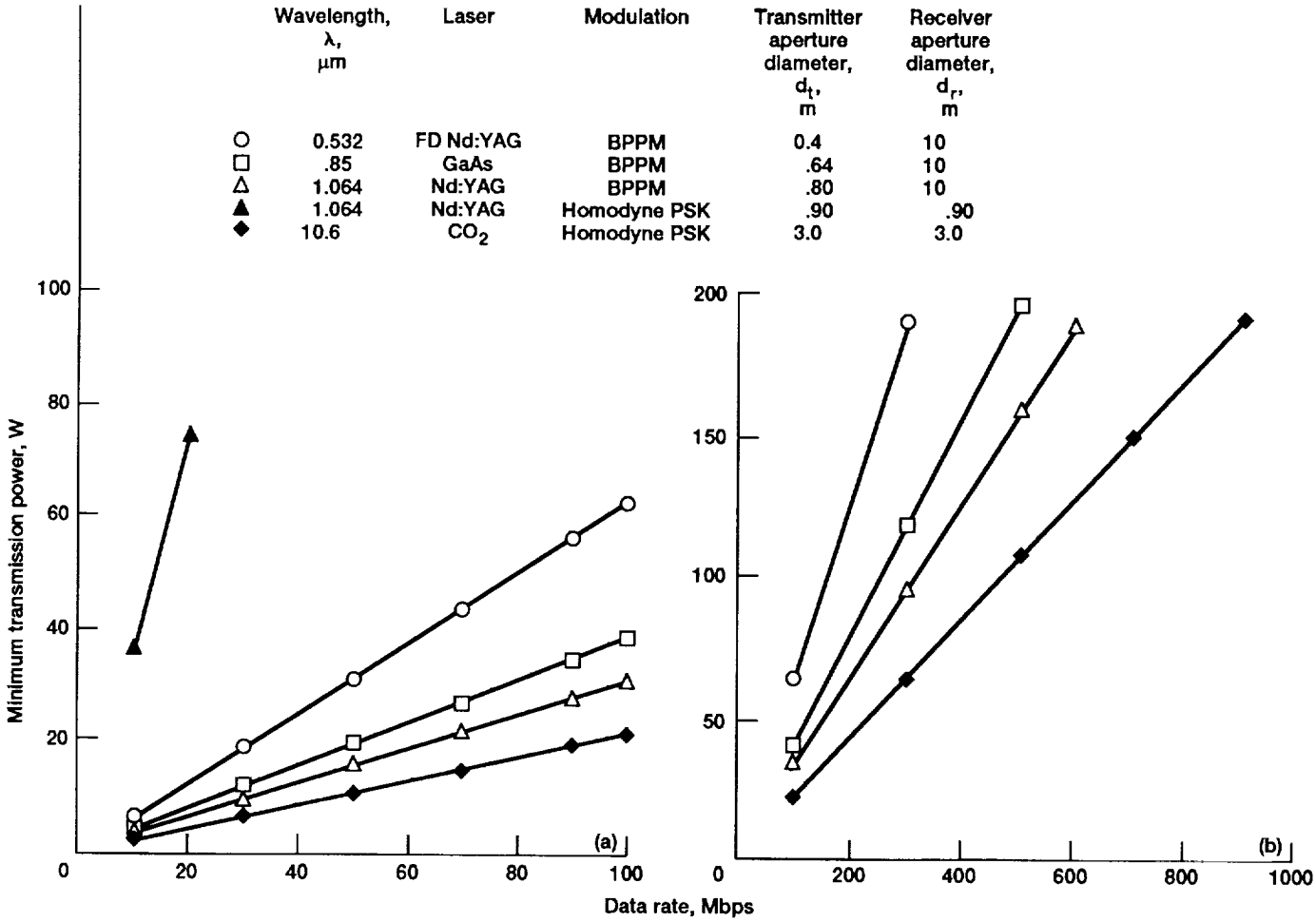
(a) 10 to 100 Mbps.
(b) 100 to 1000 Mbps.

Figure 7.4.—Transmission power versus bit rate for various optical schemes in the presence of σ = 0.05-μrad rms pointing and tracking error for optimum aperture sizes (10^{-6} BER).



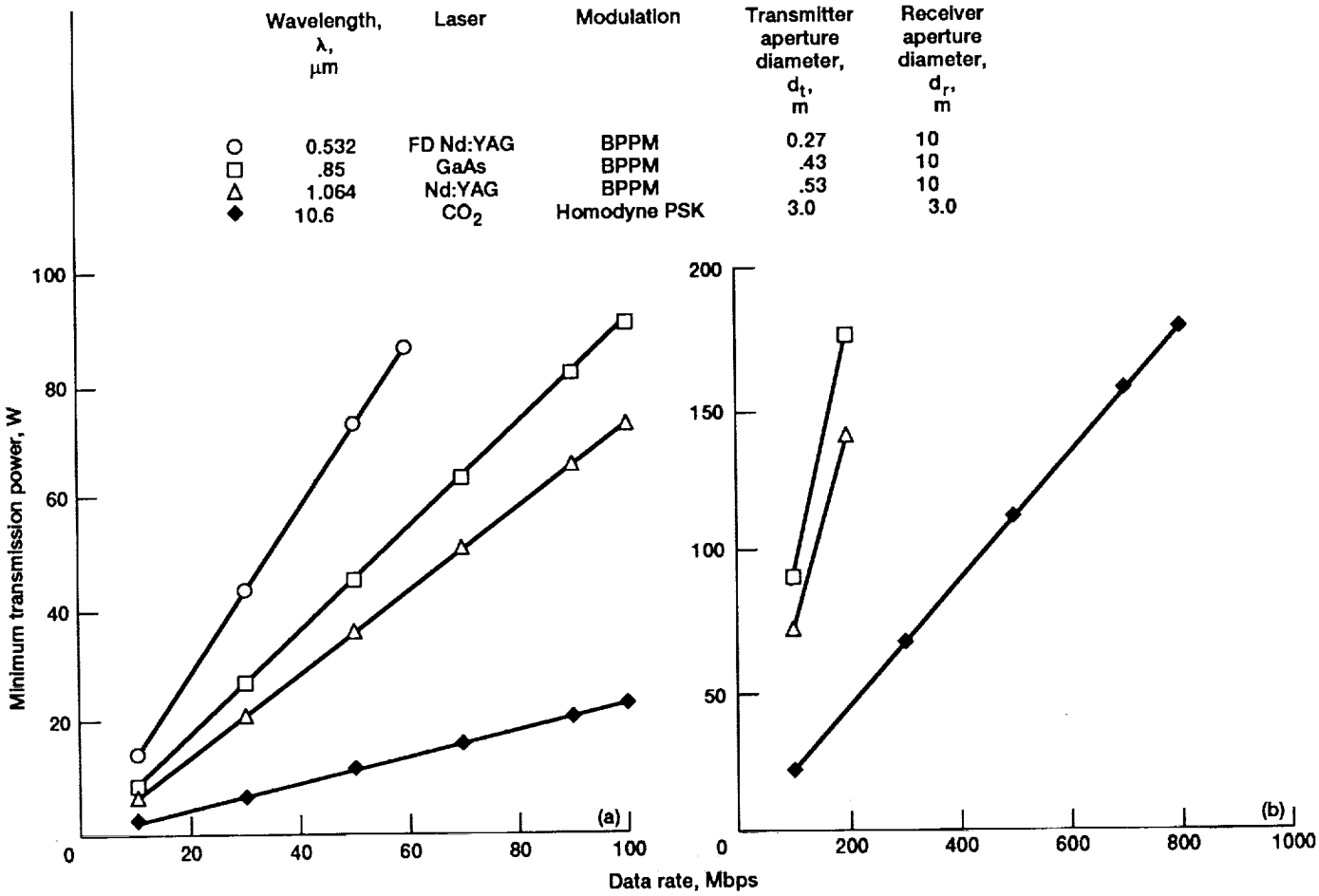
(a) 10 to 100 Mbps.
(b) 100 to 1000 Mbps.

Figure 7.5.—Transmission power versus bit rate for various optical schemes in the presence of $\sigma = 0.10$-μrad rms pointing and tracking error for optimum aperture sizes (10^{-6} BER).



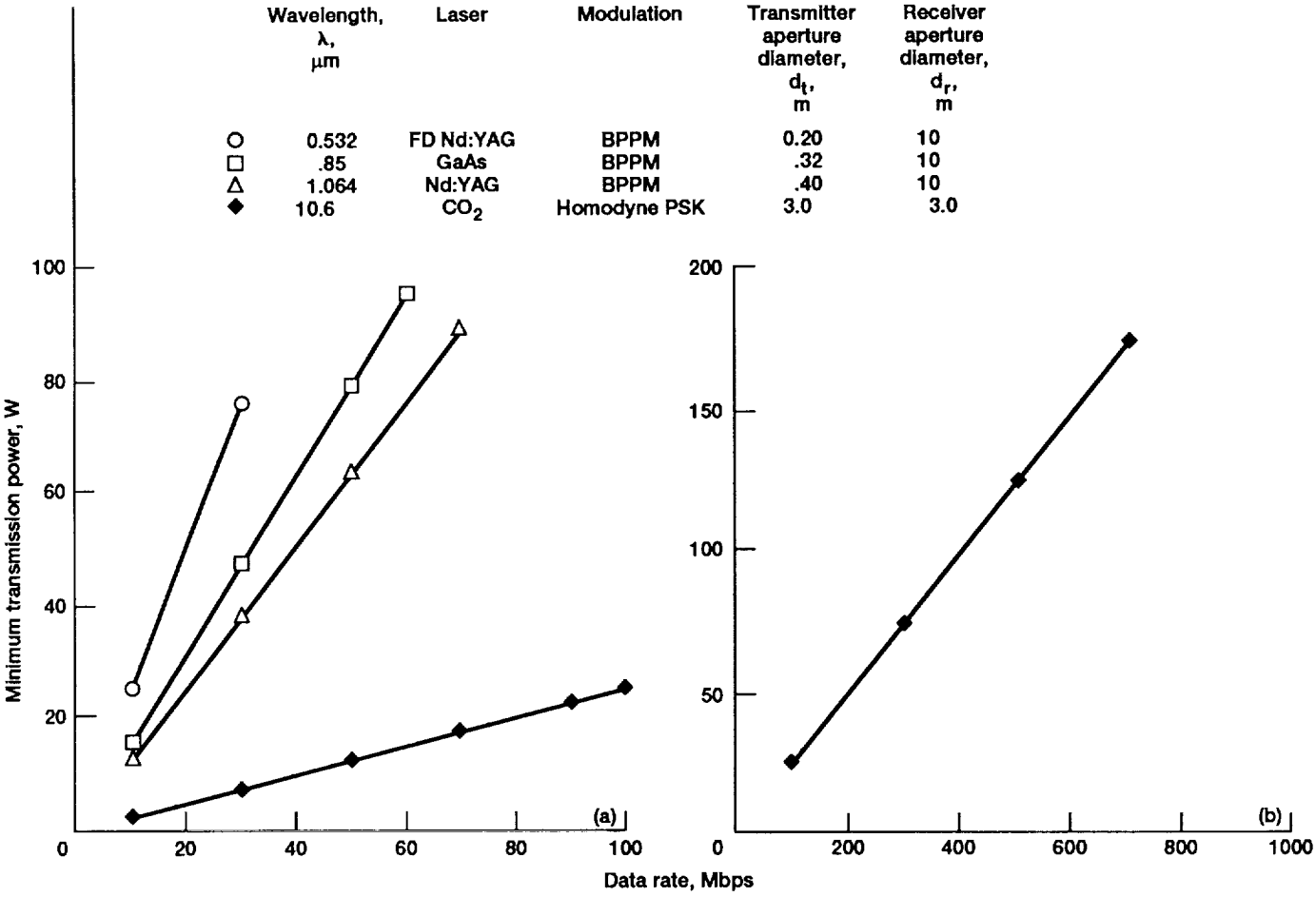
(a) 10 to 100 Mbps.
(b) 100 to 1000 Mbps.

Figure 7.6.—Transmission power versus bit rate for various optical schemes in the presence of $\sigma = 0.20$-μrad rms pointing and tracking error for optimum aperture sizes (10^{-6} BER).



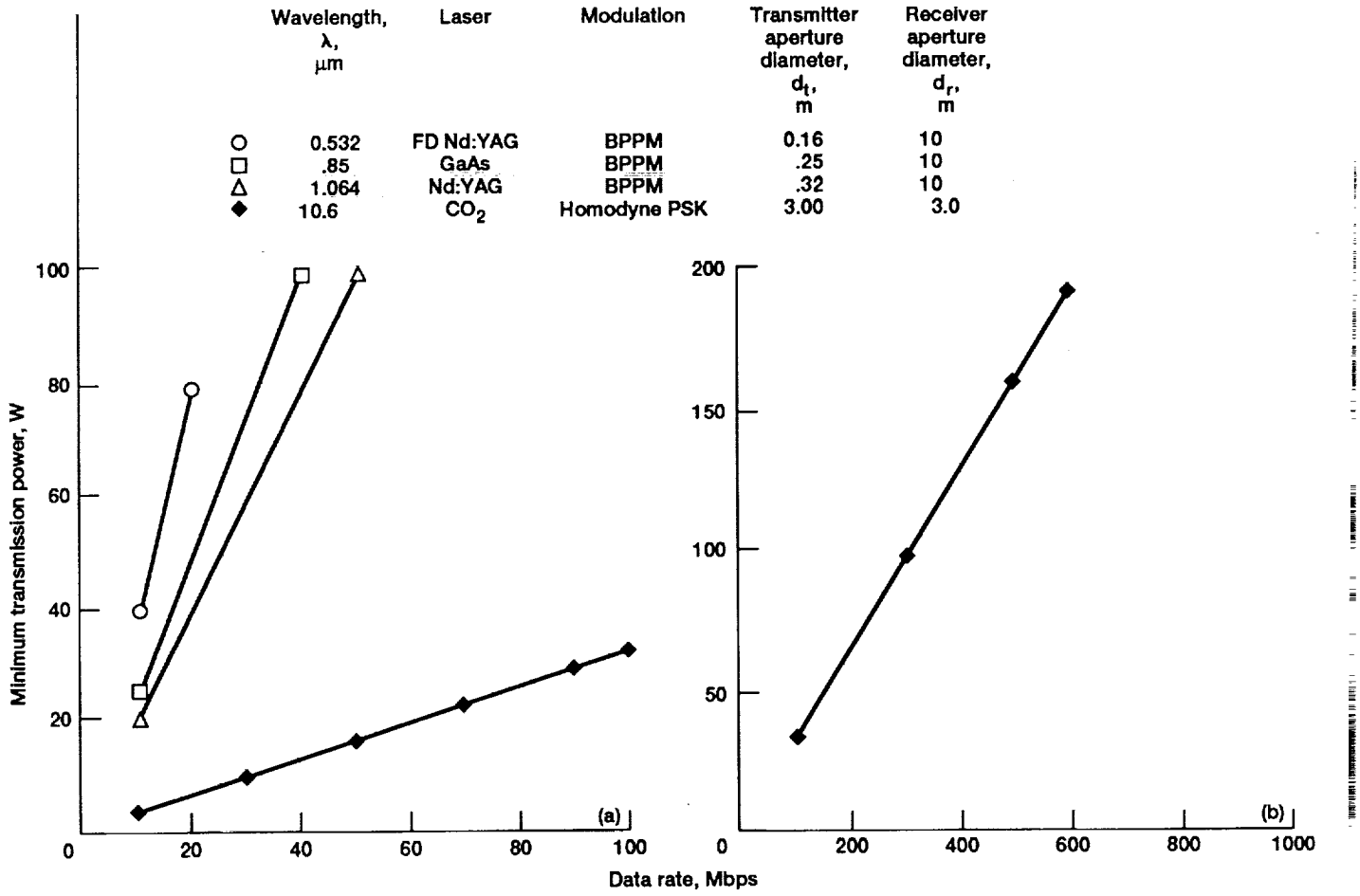
(a) 10 to 100 Mbps.
(b) 100 to 1000 Mbps.

Figure 7.7.—Transmission power versus bit rate for various optical schemes in the presence of $\sigma = 0.30$-μrad rms pointing and tracking error for optimum aperture sizes (10^{-6} BER).



(a) 10 to 100 Mbps.
(b) 100 to 1000 Mbps.

Figure 7.8.—Transmission power versus bit rate for various optical schemes in the presence of $\sigma = 0.40$-μrad rms pointing and tracking error for optimum aperture sizes (10^{-6} BER).



(a) 10 to 100 Mbps.
(b) 100 to 1000 Mbps.

Figure 7.9.—Transmission power versus bit rate for various optical schemes in the presence of $\sigma = 0.50$-μrad rms pointing and tracking error for optimum aperture sizes (10^{-6} BER).

Tables 7.6 and 7.7 summarize the results for a 100-Mbps Mars-to-Earth optical link. Table 7.6 shows the minimum power requirements for the three PPM direct-detection systems, whereas table 7.7 gives the results for the four coherent detection systems. Since the minimum-power to data-rate relationship is linear, the results for 10-Mbps and 1-Gbps links are 0.10 and 10 times the values in these tables, respectively. In comparing tables 7.6 and 7.7, we can observe that the power requirements for the coherent systems increase much more rapidly with rms error than do those for the direct-detection systems. This rapid increase results from combined effects of the transmitter pointing losses and receiver LO tracking losses to which the coherent systems are subject. From the expressions for p in equations (7.3) and (7.5), we can see that the minimized power is proportional to σ^4 for a coherent system, whereas it is proportional to σ^2 for a direct-detection system. Note also in table 7.7 that, because of the CO_2 systems's long wavelength and wide beamwidth, its power requirements are relatively unaffected by the rms error.

In relation to direct-detection for the Mars-to-Earth optical link, the results in table 7.6 suggest that the fundamental Nd:YAG laser ($\lambda = 1.064$ μm) system using a high-order PPM may be the preferred implementation. For example, if we use $M = 64$ and assume $\sigma = 0.4$ μrad, a 100-Mbps link can be supported with a 40-cm transmitter telescope, a 20-W Nd:YAG laser (1280 W peak power), and a 10-m receiving photon bucket. A 10-W FD Nd:YAG laser (capable of 640-W peak power) could also be used, but would require a more stringent pointing accuracy of $\sigma = 0.2$ μrad. The pulse repetition rate of the laser in both these cases would be 17 MHz ($PRF = R/\log_2 M$), and the pulse duration 1 ns ($\tau = \log_2 M/(MR)$). Since the power requirements for a 1-Gbps link are 10 times those shown in table 7.6, returning a 1-Gbps data stream would require a larger transmitting telescope and smaller rms pointing jitter. For example, if the fundamental Nd:YAG laser (M = 64) were used, a 1-Gbps link would require a 160-cm telescope, 14 W of average power (or 896 W of peak power), a 10-m photon bucket, and an rms pointing error of only 0.1 μrad. Note that

TABLE 7.6.—MINIMAL POWER REQUIREMENTS FOR 100-Mbps EARTH-TO-MARS PPM DIRECT-DETECTION LINK USING OPTIMAL TRANSMITTER APERTURE SIZES AND 10-m RECEIVING PHOTON BUCKET

(a) M-ary pulse-position-modulation (PPM) direction-detection (DD) link using FD Nd:YAG laser; wavelength, λ 0.532 μm.

Modulation, M	Minimal power requirements, W						
	rms pointing error, σ, μrad						
	0.05	0.1	0.15	0.2	0.3	0.4	0.5
	Half-power BW, $\theta = \lambda/D$, μrad						
	0.33	0.66	1.0	1.33	2.0	2.66	3.3
	Optimum transmitter aperture size, cm						
	160	80	53	40	27	20	16
2	4.0	15.9	35.8	63.6	143.1	254.4	397.5
4	2.1	8.3	18.6	33.1	74.5	132.4	206.8
8	1.4	5.5	12.5	22.1	49.8	88.6	138.4
16	1.0	4.1	9.3	16.5	37.2	66.1	103.3
32	.8	3.3	7.4	13.2	29.7	52.7	82.4
64	.7	2.7	6.2	11.0	24.7	43.9	68.5
128	.6	2.4	5.3	9.4	21.2	37.7	58.9
256	.5	2.1	4.7	8.3	18.7	33.2	51.9

(b) M-ary PPM DD link using Nd:YAG laser; wavelength, λ, 1.064 μm.

Modulation, M	Minimal power requirements, W						
	rms pointing error, σ, μrad						
	0.05	0.1	0.15	0.2	0.3	0.4	0.5
	Half-power BW, $\theta = \lambda/D$, μrad						
	0.33	0.66	1.0	1.33	2.0	2.66	3.3
	Optimum transmitter aperture size, cm						
	320	160	106	80	54	40	32
2	2.0	8.0	17.9	31.8	71.5	127.2	198.8
4	1.0	4.1	9.3	16.5	37.2	66.2	103.4
8	.7	2.8	6.2	11.1	24.9	44.3	69.2
16	.5	2.1	4.6	8.3	18.6	33.1	51.6
32	.4	1.6	3.7	6.6	14.8	26.4	41.2
64	.3	1.4	3.1	5.5	12.3	21.9	34.3
128	.3	1.2	2.7	4.7	10.6	18.9	29.5
256	.2	1.0	2.3	4.2	9.3	16.6	25.9

(c) M-ary PPM DD link using GaAs laser; wavelength, λ, 0.85 μm.

Modulation, M	Minimal power requirements, W						
	rms pointing error, σ, μrad						
	0.05	0.1	0.15	0.2	0.3	0.4	0.5
	Half-power BW, $\theta = \lambda/D$, μrad						
	0.33	0.66	1.0	1.33	2.0	2.66	3.3
	Optimum transmitter aperture size, cm						
	256	128	85	64	43	32	26
2	2.5	10.0	22.4	39.8	89.6	159.2	248.8
4	1.3	5.2	11.6	20.7	46.6	82.8	129.4
8	.9	3.5	7.8	13.9	31.2	55.4	86.6
16	.6	2.6	5.8	10.3	23.3	41.4	64.6
32	.5	2.1	4.6	8.3	18.6	33.0	51.6
64	.4	1.7	3.9	6.9	15.4	27.4	42.9
128	.4	1.5	3.3	5.9	13.3	23.6	36.9
256	.3	1.3	2.9	5.2	11.7	20.8	32.5

this pointing accuracy would have to be achieved in an open-loop mode (i.e., without the aid of a tracking beacon from the receiver) because of the long signal-propagation delay. Finally, the results in table 7.6 for the GaAs laser indicate that the power requirements are substantially beyond the power output levels anticipated for semiconductor lasers in the foreseeable future; output power from semiconductor devices is projected to be only on the order of 1 W. In addition, because semiconductor lasers have a low peak-to-average-power ratio, they are not suitable for high-order PPM. The semiconductor laser diodes' limited output power capability would necessitate the use of a low-order PPM ($M = 2$ or 4), an extremely small rms error (<0.1 μrad), and a fairly large aperture size (>2.5 m). Hence, the application of semiconductor lasers for the high-data-rate Mars-to-Earth optical communication link appears impractical.

The minimum power requirements for the various coherent detection links at 100 Mbps are shown in table 7.7. Since the rms pointing error and rms LO tracking error are assumed to be equal for these systems, the optimum transmitter and receiver aperture sizes also are considered to be equal. It is obvious from table 7.7 that the coherent systems are much

TABLE 7.7.—MINIMUM POWER REQUIREMENTS FOR 100-Mbps EARTH-TO-MARS COHERENT DETECTION LINKS USING OPTIMUM TRANSMITTER AND RECEIVER APERTURE SIZES

(a) Homodyne PSK link using FD Nd:YAG laser; wavelength, λ, 0.532 μm.

	rms pointing error, σ, μrad						
	0.05	0.1	0.15	0.2	0.3	0.4	0.5
Optimum transmitter and receiver aperture size, cm	180	90	60	45	30	23	18
Half-power BW, $\theta = \lambda/D$, μrad	0.30	0.60	0.89	1.18	1.77	2.31	2.9
Transmitter power, W	11.8	188.4	953.9	3015	15 263	48 238	117 769

(b) NCFSK link using GaAs diode laser; wavelength, λ, 0.85 μm.

	rms pointing error, σ, μrad						
	0.05	0.1	0.15	0.2	0.3	0.4	0.5
Optimum transmitter and receiver aperture size, cm	280	140	93	70	47	35	28
Half-power BW, $\theta = \lambda/D$, μrad	0.30	0.61	0.91	1.21	1.81	2.43	3.0
Transmitter power, W	15.2	243.0	1230	3887	19 678	62 193	151 839

(c) Homodyne PSK link using Nd:YAG laser; wavelength, λ, 1.064 μm.

	rms pointing error, σ, μrad						
	0.05	0.1	0.15	0.2	0.3	0.4	0.5
Optimum transmitter and receiver aperture size, cm	360	180	120	90	60	45	36
Half-power BW, $\theta = \lambda/D$, μrad	0.30	0.60	0.89	1.18	1.77	2.31	2.9
Transmitter power, W	1.5	23.6	119.2	376.9	1908	6030	14 721

(d) Homodyne PSK link using CO_2 laser; wavelength, λ, 10.6 μm.

	rms pointing error, σ, μrad						
	0.05	0.1	0.15	0.2	0.3	0.4	0.5
Optimum transmitter and receiver aperture size, cm	300	300	300	300	300	300	300
Half-power BW, $\theta = \lambda/D$, μrad	3.5	3.5	3.5	3.5	3.5	3.5	3.5
Transmitter power, W	20.8	20.9	21.2	21.5	22.7	25.4	32.2

Power values for R = 10 Mbps and for R = 1 Gbps are equal to 0.1 and 10 times these values, respectively.

[a]The CO_2 laser system uses suboptimal 3-m apertures. Optimal sizes yielding minimal requirements range in diameter from 3.6 to 36 m.

more sensitive to the spatial pointing and tracking error than are the direct-detection systems (with the exception of the CO_2 system). The coherent system's detection sensitivity advantage (11 dB for homodyne PSK) is only useful when the rms error can be made extremely small. Given that the projected output power of a coherent Nd:YAG laser is 10 W (1.064 μm), a 100-Mbps link will require an rms pointing and LO tracking error of less than 0.1 μrad. Table 7.7 shows that for an rms error of 0.05 μrad, 100 Mbps can be supported with a 12-W FD Nd:YAG laser using 1.8-m transmitting and receiving telescopes or with a 2-W Nd:YAG laser using 3.6-m telescopes. Achievement of 0.05-μrad accuracy would probably require a very costly and complex control system. (By comparison, the Hubble Space Telescope has an rms error of 0.034 μrad.)

The long wavelength of the CO_2 laser and its wide beamwidth allow this system to operate with a much more relaxed pointing tolerance. As indicated in table 7.7, a 100-Mbps link at $\sigma = 0.5$ μrad can be implemented with 3-m telescopes and a 32-W laser. This power output should be achievable with future CO_2 waveguide lasers. In addition, relaxed surface tolerances resulting from the long wavelength should simplify development of the relatively large optics. Selection of the CO_2 system would require the development of a space-qualified CO_2 laser with high power output and long life. Unfortunately, research and development work on the CO_2 laser for space communication applications is limited at present. As mentioned in section 7.2.1.3, however, the ESA has been developing and experimenting with a CO_2 PSK homodyne transceiver for the past several years (ref. 23).

9.0 Conclusions

In conclusion, the performance of the optical pointing and tracking subsystem will in large part determine what type of optical system is best suited for supporting high-data-rate Mars-to-Earth communications. The presence of random and static spatial pointing and tracking errors constrains the maximum telescope size that can be used, and therefore places minimum power requirements on the laser transmitter. Because of its relative simplicity, tolerance to mispointing, and high-power output capability, a Nd:YAG laser direct-detection system using a high-order PPM (pulse position modulation) appears to be the most feasible candidate for the long-range Mars-to-Earth optical link. A diode laser system may also become feasible if a long-life, high-power, multiwatt laser diode array producing a near diffraction-limited beam can be developed. The CO_2 system. although it might have a significant power advantage, appears unlikely because of its complexity and unreliability.

Lewis Research Center
National Aeronautics and Space Administration
Cleveland, Ohio, November 8, 1991

Appendix A
Symbols

A_{IF}	IF carrier amplitude
A	electric field amplitude, V/m
A_d	area of circular detector
$A_i(t)$	amplitude of ith symbol in ASK waveform set
A_L	LO field amplitude
A_o	amplitude of the incident field on the aperture
A_r	receiver aperture area
A_S	signal field amplitude
A_s	circular disc area of background source
A_v	VCO proportionality constant
AM	amplitude modulation
APD	avalanche photodiode
ASK	amplitude shift keying
ATS-F	Advanced Technology Satellite—Spacecraft F
ATS-G	Advanced Technology Satellite—Spacecraft G
AU	astronomical units
a	receiver aperture radius
B	bandwidth
B_{IF}	IF bandwidth
B_L	single-sided noise bandwidth of ideal low-pass filter
B_m	information-signal bandwidth
B_N	noise-equivalent single-sided bandwidth
$B_{N_{opt}}$	optimum value of the loop bandwidth
BER	instantaneous bit error rate
BER_{floor}	irreducible error rate
BER_ψ	average BER due to laser phase noise
$\overline{BER}$	time-averaged BER
BPPM	binary pulse position modulation
BPSK	binary phase shift keying
BW	beam width
b	pointing bias
C	total intrinsic capacitance of the detector-preamplifier circuit
C_A	APD capacitance
C_Q	preamplifier capacitance
CLEO	Conference on Lasers and Electro-Optics
CW	continuous wave
c	speed of light in vacuum, 2.998×10^8 m/sec
$c_{\tau_{coh}}$	coherence length
D	aperture diameter
DD	direct detection
DP	diode pumped
DPSK	differential phase shift keying
DSB/SC	double sideband carrier suppressed
d	distance between centers of two overlapping Airy discs of equal diameter
d_{coh}	diameter of a circular area over which the radiation has a high degree of spatial coherence
d_{opt}	optimum transmitter aperture diameter
d_r	receiver aperture diameter, m
d_{random}	optimum aperture diameter in the presence of random pointing error
d_{static}	optimum aperture diameter in the presence of static pointing error
d_t	transmitter aperture diameter, m
dc	direct current
$d\lambda$	wavelength interval
$d\nu$	frequency interval
E	symbol energy
E_b	bit energy
E_L	LO electric field amplitude
E_{Rc}, E_{Lc}	field patterns exactly centered on each other
E_S	signal electric field amplitude
ERS	Earth relay satellite
ESA	European Space Agency
$E(r)$	diffracted optical field in the detector plane
$E(t)$	electric field of unmodulated optical carrier signal
$\overline{E(t)^2}$	mean-square value of electric field amplitude
E-O	electric-optic
e	electron charge, 1.6×10^{-19} C
F	excess noise factor, ratio of the mean-squared gain to the mean gain squared
FD	frequency doubled
FEC	forward error correction
FK_N	signal-independent noise component
FK_S	signal-dependent noise component
FM	frequency modulation
FOV	field of view
FSK	frequency shift keying
$F(s)$	transfer function of loop filter in phased-locked loop

f	focal length
f_L	LO frequency
f_o	center frequency
f_S	optical carrier frequency, Hz
G	APD random gain
$\bar{G}$	APD mean gain
GEO	geosynchronous Earth orbit
GI_S	photomultiplied signal current
$G(s)$	open-loop transfer function
$G_x(f)$	two-sided power spectral density of the noise in the detector-preamplifier output current
$G_y(f)$	output-noise power spectral density
$(\bar{G})^2$	squared mean of the APD gain
$H(f)$	closed-loop transfer function of the spatial tracking loop
$H(s)$	closed-loop transfer function
$H(\nu)$	spectral irradiance
h	Planck constant, 6.624×10^{-34} J/sec
h_d	detector height along the y-axis
$h(t)$	impulse response
$h\nu$	photon energy, J/photon, hc/λ
I	electric field intensity
I_o	field intensity
$\bar{I}$	average detector current
I_B	average background current
I_D	average detector dark current
I_{DB}	APD bulk dark current
I_{DS}	APD surface dark current
I_{MAX}	maximum on-axis intensity in the focal plane
$I_o(z)$	modified Bessel function of the first kind of order zero
I_S	average signal current, $\alpha e P_S$
I_s	average received signal intensity
IF	intermediate frequency
IM	intensity modulated
IPS	Instrument Pointing System
$I(r)$	intensity distribution in the focal plane
ISL	intersatellite link
$I(t)$	intensity of an intensity-modulated carrier
$I(\nu)$	spectral radiance
$i_d(t)$	component of photodetector output current resulting from the desired signal
$i_{IF}(t)$	desired signal current
$J_1(x)$	first-order Bessel function of the first kind
j	imaginary unit
K	defined amplitude constant
K_B	noise photoelectron counts caused by background radiation
K_d	phase detector constant
K_{DB}	noise photoelectron counts caused by bulk dark current
K_{DS}	noise photoelectron counts caused by surface dark current
K_S	average number of photoelectrons per symbols, or photoelectrons per bit when $M = 2$
K_T	thermal noise count
K_v	VCO constant
k	Boltzmann constant, 1.381×10^{-23} J/K
k_b	number of bits
k_p	propagation number $(2\pi/\lambda)$
k_{eff}	ratio between the hole and electron-ionization coefficients, or the effective ionization ratio
k_L	LO field propagation vector $(k_L = k_x i + k_y j + k_z k)$
k_S	signal field propagation vector $(k_S = k_x i + k_y j + k_z k)$
L_o	other losses (e.g., obscuration, implementation, filter)
L_p	power penalty
L_t	transmitter pointing loss
LBO	lithium triborate
LDR	large deployable reflector
$L(f)$	Lorentzian function
LITE	Laser Intersatellite Transmission Experiment
LO	local oscillator
LOS	receiving line of sight
M	number of PPM time slots
M-ary	waveform t with M possible symbols
M_{IM}	intensity modulation index
MIT-LL	Massachusetts Institute of Technology—Lincoln Laboratories
MQW-APD	multiple-quantum-well avalanche photodiode
MRS	Mars relay satellite
m_n	mean of count distribution (gaussian) for the nonsignal PPM time slot at the integrator output
m_s	mean of count distribution (gaussian) for the signal PPM time slot at the integrator output

$m(t)$	information signal
N_o	one-sided power spectral density of zero-mean additive, white gaussian thermal noise
N_S	one-sided power spectral density of shot noise
$N_S/2$	two-sided power spectral density of LO shot noise
N_T	two-sided thermal-noise spectral density
NCFSK	noncoherent phase shift keying
$n_s(t)$	LO shot noise signal
OOK	on-off keying
P	average power
P_a	average laser power
P_B	average received background power
$P_E(K_S)$	bit error rate as function of the received signal counts per slot
P_{IF}	IF power
P_{ideal}	required power with zero pointing error
P_L	LO power
P_m	average power of $m(t)$
$P_{\min}$	minimum transmission power
P_o	incident power
P_p	peak power of optical pulse
P_r	received average power
P_S	average received signal power
P_t	average transmitter laser power, W
PCM	pulse code modulation
PDF	probability density function
PIN	P-type, Intrinsic, N-type
PLL	phase-locked loop
PM	phase modulation
PMT	photomultiplier tube
PPM	pulse position modulation
PRF	pulse repetition frequency
PSD	power spectral density
PSK	phase shift keying
Prob($n < s$)	probability that the photoelectron count "n" in an arbitrary nonsignal slot is less than "s"
PWC	probability of a correct PPM word
PWE	probability of a PPM word error
p	pointing parameter in direct-detection system, $G_t\sigma_T^2$ (*p.* 74)
$p_n(x)$	gaussian probability density of photoelectron count for nonsignal time slot
p_r	receiver pointing parameter in coherent detection system, $G_r\sigma_R^2$
$p_{r_{\text{opt}}}$	optimum pointing parameter
$p_s(t)$	instantaneous received signal power
$p_s(x)$	gaussian probability density of photoelectron count for signal time slot
p_t	transmitter pointing parameter in coherent detection system, $G_t\sigma_T^2$
$p_{t_{\text{opt}}}$	optimum transmitter pointing parameter
$p(\theta_t)$	probability density function of radial pointing and tracking error
$p(\theta_t,\theta_r)$	joint PDF of transmitter pointing error and LO tracking error
Q	quality factor
$Q_M(a,b)$	Marcum function
QAM	quadrature amplitude modulated
QE	quantum efficiency
QPPM	quaternary pulse position modulation
q	pointing parameter, $\alpha P_t/\sigma_T^2$
q_{opt}	optimum pointing parameter
R	bit rate; radius of Airy disc
R_b	channel data rate
R_L	detector load resistance
R_S	symbol rate
RC	resistor capacitor
RF	radiofrequency
r	arbitrary position vector $(xi + yj + zk)$
r_d	radius of circular detector
r_L	$(x - x_L)^2 + (y - y_L)^2$
r_R	$(x - x_R)^2 + (y - y_R)^2$
rms	root mean square
$r(t)$	received signal at the filter output
S	average signal power
SAM-APD	separate-absorption-and-multiplication-region avalanche photodiode
SEI	Space Exploration Initiative
$S(f)$	tracking error
$S_D(f)$	single-sided PSD of the tracking detector noise at the input of the spatial tracking loop
$S_p(f)$	single-sided PSD of the platform error at the input of the spatial tracking loop
$S_\psi(f)$	one-sided phase noise power spectral density
$S_\omega(f)$	one-sided frequency noise power spectral density
$S_{\psi\text{IF}}(f)$	one-sided power spectral density of the IF signal phase noise
SNR	signal-to-noise ratio

SNR_D	demodulated *SNR* for a strong laser local oscillator
s	Laplace transform variable
$s(t)$	phased-locked loop input signal
$s_d(t)$	phase detector output
$s_i(t)$	transmitted signals
$s_o(t)$	output signal from the voltage controlled oscillator
s_v	VCO input signal
T	effective noise temperature of the detector preamplifier
T_s	symbol period
T_b	bit period
T_{eq}	preamplifier noise equivalent temperature
T_{sun}	temperature of black-body radiator approximating Sun's temperature, 6000 K
TEM_{00}	transverse electromagnetic mode
t	time
u	normalized pointing error, θ_t/σ_T
VCO	voltage controlled oscillator
v	normalized tracking error, θ_r/σ_R
W_f	detector-preamplifier circuit bandwidth
W_P	pulse bandwidth
w	detector width along the x-axis
X	random variable, denotes photoelectron count at the integrator output
x,y,z	coordinates
x_c	axial coordinate of the center of the focused field
x_L	$f - \tan(\theta_{yR})$
x_R	$\text{ftan}(\theta_{xR}) = f - \tan(\theta_{yR})$
y_c	transverse coordinate of the center of the focused field
$y(t)$	demodulated output signal
Z_o	free-space impedance, $\approx 377\omega$
Z	link range
α	conversion factor that converts average received power (in watts) into the average rate of photoelectron emission from the detector (in photoelectrons per sec), $\eta/h\nu$
α_t	parameter independent of θ_t, P_t, and d_t
β	FM modulation index ($\Delta f/B_m$)
γ	random frequency noise
$\Delta\lambda$	laser line width, mm
$\Delta\nu$	bandwidth of the radiation
Δf	optical bandwidth
ς	damping factor
η	photodetector quantum efficiency, (number of emitted photoelectrons/incident photon)
η_{Fa}	alignment efficiency
η_{het}	overall heterodyning efficiency, or SNR degradation
η_{pol}	polarization efficiency
η_r	receiver optics efficiency
η_t	transmitter optics efficiency
$\Theta(s)$	Laplace transform of $\Theta(t)$
$\Theta(t)$	phase of the received signal
θ	optical carrier phase
θ_{dL}	diffraction-limited planar angle field of view
θ_{fv}	receiver planar-angle field of view
θ_{IF}	$\theta_S - \theta_L$
θ_L	LO initial phase
θ_o	angular radius of Airy disc
θ_S	signal initial phase
θ_t	instantaneous radial pointing and tracking error
θ_x,θ_y	azimuth and elevation angles
θ_{xL},θ_{yL}	angles by which LO field is off-axis from normal
θ_{xR},θ_{yR}	angles by which received field is off-axis from normal
λ	optical wavelength, m
λ_L	wavelength of LO field
λ_o	center wavelength
λ_R	wavelength of received field
μ_n	mean value of integrator output during nonsignal slot times
μ_s	mean value of integrator output during signal slot times
ρ	multiplying constant for phase-error variance
ρ_b	radius of circular background source
σ	rms pointing and tracking error
σ_ψ^2	total phase-error variance
$\sigma_{\psi\text{min}}{}^2$	minimum phase-error variance
σ_G^2	variance of gain (measure of randomness of APD gain)
σ_n^2	variance in photoelectron count during a nonsignal slot
σ_{PN}^2	phase error variance resulting from laser phase noise
σ_R	rms receiver pointing error

σ_s^2	variance in photoelectron count during a signal slot
σ_{SN}^2	phase error variance resulting from LO shot noise
σ_T	rms transmitter pointing error
σ_x^2	variance
τ	PPM time slot width, sec
τ_{coh}	coherence time
τ_p	width of optical pulse, sec
ν	optical frequency, Hz
ν_d	FSK tone spacing
$\Phi(s)$	Laplace transform phase error
φ	angle between polarization vectors $\vec{A}_S$ and $\vec{A}_L$ (eq. (2.35))
$\phi_i(t)$	phase of the ith symbol in M-PSK symbol set
$\Psi(s)$	phase error transfer function
$\psi(t)$	phase error
$\psi(x,y,z,t)$	wave function, or E-field
Ω	solid angle describing conical field of view
Ω_{dL}	diffraction-limited solid angle field of view
Ω_{fv}	receiver solid angle field of view
Ω_r	solid angle subtended by the receiver aperture when viewed from the source
Ω_s	solid angle subtended by the source when viewed from the receiver
ω_{IF}	radian frequency of IF signal, $\omega_S - \omega_L$
ω_L	LO optical frequencies, rad/sec
ω_n	natural frequency
ω_o	center frequency, rad/sec
ω_S	optical carrier frequency, rad/sec
$\mathcal{L}$	Laplace transform
$\nabla^2\psi$	free-space differential wave equation
$4kT/R_L$	one-sided power spectral density of thermal noise

Appendix B
Fraunhofer Diffraction by a Circular Aperture

Here we will use the Fraunhofer diffraction integral, which is essentially a two-dimensional Fourier transform, to derive the far-field intensity pattern of a uniformly illuminated circular aperture (i.e., an approximation of the far-field pattern of a diffracted laser beam transmitted through a telescope's circular aperture). The Fraunhofer integral also may be used to compute the diffracted field in the focal plane of a lens, given the field distribution over the front of the lens.

Consider a monochromatic wave (not necessarily a plane wave) propagating along the z-axis, whose plane of origin, or source plane, is characterized by the E-field distribution ($E(\tilde{x},\tilde{y},0)$). We want to determine the E-field distribution $E(x,y,z)$ on the observation plane at a distance of z away. In the case of a laser free-space link, the source plane would correspond to the transmitter telescope aperture plane, and the observation plane would correspond to the receiver telescope aperture plane (z is a very large distance). In the case of a receiving lens, the source plane would be the plane at the front surface of the lens, and the observation plane would be the focal plane of the lens (z is the focal length). The field at the observation plane is related to the field on the source plane through the Fraunhofer diffraction integral, which is

$$E(x,y,z) = -\frac{je^{jkz}}{\lambda z} e^{jk(x^2+y^2)/2z} \int\int E(\tilde{x},\tilde{y},0)\, e^{-jk(x\tilde{x}+y\tilde{y})/z} d\tilde{x}\, d\tilde{y} \tag{B.1}$$

If we now consider a circular aperture of diameter $D = 2a$ upon which is incident a uniform amplitude monochromatic plane wave (i.e, $E(\tilde{x}, \tilde{y}, 0) = E_o$), then the double integral portion in equation (B.1) becomes

$$E_o \int\int e^{-jk(x\tilde{x}+y\tilde{y})/z}\, d\tilde{x}\, d\tilde{y} \tag{B.2}$$

Converting to polar coordinates, we have

$$x = r\cos\theta \quad y = r\sin\theta$$

$$\tilde{x} = \tilde{r}\cos\tilde{\theta} \quad \tilde{y} = \tilde{r}\sin\tilde{\theta}$$

$$x\tilde{x} + y\tilde{y} = r\tilde{r}(\cos\theta\cos\tilde{\theta} + \sin\theta\sin\tilde{\theta}) \tag{B.3}$$

$$= r\tilde{r}\cos(\tilde{\theta} - \theta)$$

$$d\tilde{x}\, d\tilde{y} = d\tilde{A} = \tilde{r}\, d\tilde{r}\, d\tilde{\theta}$$

The integral in equation (B.2) is therefore

$$\int\int e^{-jk(x\tilde{x}+y\tilde{y})/z}\, d\tilde{x}\, d\tilde{y} = \int_0^{2\pi}\int_0^{a} e^{-jkr\tilde{r}\cos(\tilde{\theta}-\theta)/z}\, \tilde{r}\, d\tilde{r}\, d\tilde{\theta} \tag{B.4}$$

Now consulting standard integral tables, we find that

$$\int_0^{2\pi} e^{-jkr\tilde{r}\cos(\tilde{\theta}-\theta)/z}\, d\tilde{\theta} = 2\pi J_o\left(\frac{kr\tilde{r}}{z}\right) \tag{B.5}$$

and that

$$\int_0^{a} J_o\left(\frac{kr\tilde{r}}{z}\right)\tilde{r}\, d\tilde{r} = \frac{a^2}{2}\frac{2J_1(kar/z)}{kar/z} \tag{B.6}$$

where $J_o(x)$ is the zeroth-order Bessel function of the first kind and $J_1(x)$ is the first-order Bessel function of the first kind. Using these results, the electric field in the observation plane is

$$E(r,z) = \left(-je^{jkz}e^{jkr^2/2z}\right)\frac{\pi a^2}{\lambda z} E_o \frac{2J_1(kar/z)}{(kar/z)} \tag{B.7}$$

To find the intensity distribution in the observation plane, we take the magnitude of (B.7), square it, and divide by $2Z_o$, where Z_o is the impedance of the surrounding medium. This yields

$$I(r,z) = \frac{E_o^2}{2Z_o}\left(\frac{\pi a^2}{\lambda z}\right)^2\left[\frac{2J_1(kar/z)}{kar/z}\right]^2 \tag{B.8}$$

Since $I_o = E_o^2/2Z_o$ is the intensity of the field in the source plane, and the bracketed term involving the Bessel function has a maximum value of one at the center of the pattern ($r = 0$), the intensity distribution may be expressed as

$$I(r,z) = I_{\max}\left[\frac{2J_1(kar/z)}{(kar/z)}\right]^2 \tag{B.9}$$

where

$$I_{\max} = I_o(\pi a^2/\lambda z)^2$$

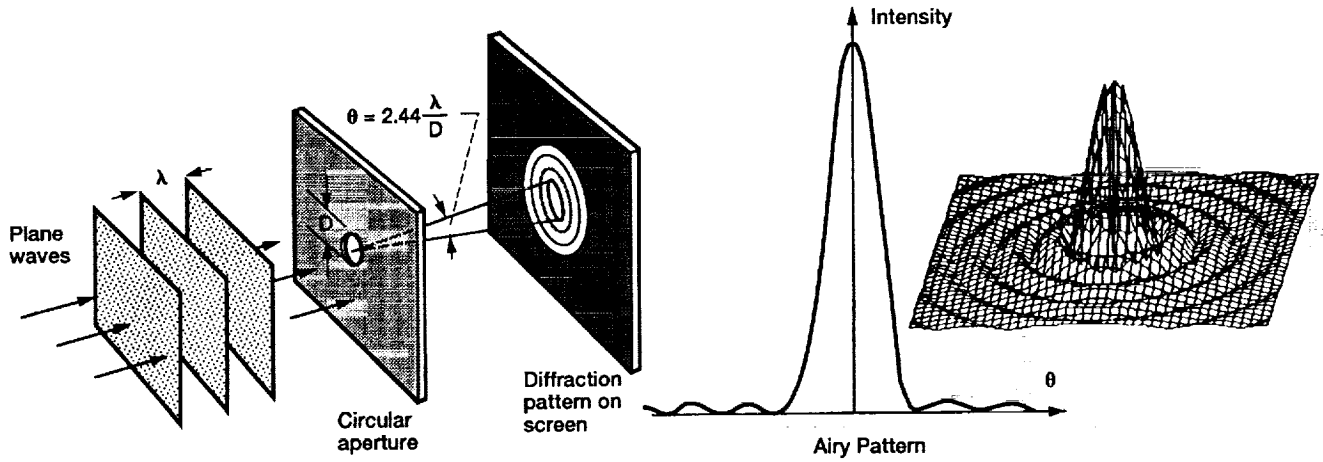


Figure B1.—Plane wave diffraction of light by a circular aperture.

A plot of equation (B.9) is shown in figure B.1. Note that the pattern is circularly symmetric and consists of a central main lobe surrounded by concentric rings of lesser intensity. This pattern also is called the Airy diffraction pattern, after the British astronomer G.B. Airy, who derived it in 1835. The diameter of the main lobe can be found by considering the first zero of the Bessel function. Since $J_1(x) = 0$ at $x \approx 1.22\pi$, the radius of the main lobe is given by

$$r_o = \frac{1.22\pi z}{ka} = \frac{1.22\pi z\lambda}{2\pi a} = 1.22\,\frac{\lambda z}{D} \qquad \text{(B.10)}$$

where $k = 2\pi/\lambda$ has been used. Thus, the diameter of the Airy disc is $d_o = 2.44\lambda z/D$. The corresponding angular radius, or diffraction-limited divergence angle (the half-angle subtended by the main lobe from the source plane), is

$$\tan\theta_o = \frac{r_o}{z} = 1.22\,\frac{\lambda}{D} \approx \theta_o \qquad \text{(B.11)}$$

The approximation is valid since the aperture diameter D is many times larger than the optical wavelength. As we would expect, equation (B.11) indicates that the smaller the wavelength and the larger the aperture size, the smaller the divergence angle and beam spot size at the observation plane. It is interesting to compare equations (B.10) and (B.11) for optical and microwave frequencies. For example, consider a satellite at Mars transmitting a 60-GHz microwave beam back to Earth through a 5-m aperture and where the distance z is 2.5 AU (1 AU = 149 600 000 km). In this case, equation (B.11) gives a beam divergence of 1 mrad, which yields a beam spot size of 72 Earth diameters (the Earth's diameter = 12 756 km). A semiconductor laser transmitting at a wavelength of 0.85 μm through a 30-cm aperture, on the other hand, has a diffraction-limited beam divergence of 3.5 μrad and a spot size of only 0.2 Earth diameters.

References

1. Yariv, A.: Optical Electronics. 3rd ed., Holt, Rinehart, and Winston, Inc., 1985.
2. McIntyre, R.J.: The Distribution of Gains in Uniformly Multiplying Avalanche Photodiodes: Theory. IEEE Trans. Electron Devices, vol. ED-19, no. 6, June 1972, pp. 703-713.
3. Conradi, J.: The Distribution of Gains in Uniformly Multiplying Avalanche Photodiodes: Experimental. IEEE Trans. Electron Devices, vol. ED-19, June 1972, pp. 713-718.
4. Gagliardi, R.M.; and Karp, S.: Optical Communications. Wiley, 1976.
5. Milonni, P.W.; and Eberly, J.H.: Lasers. Wiley, 1988.
6. Cooper, G.; and McGillem, C.: Modern Communications and Spread Spectrum. McGraw-Hill, 1986.
7. Keiser, G.: Optical Fiber Communications. McGraw-Hill, 1983.
8. Mecherle, G.S.: Maximized Data Rate Capability for Optical Communication Using Semiconductor Devices with Pulse Position Modulation. Ph.D dissertation, University of Southern California, 1986.
9. Chen, C.C.; and Gardner, C.S.: Comparison of Direct and Heterodyne Detection Optical Intersatellite Communication Links. (UIUL-ENG-87-2548, Department of Electrical and Computer Engineering, University of Illinois; NASA Contract NSG-5049) NASA CR-180210, 1987.
10. Ziemer, R.E.; and Peterson, R.L.: Digital Communications and Spread Spectrum Systems. Macmillan, 1985.
11. Spilker, J.J.: Digital Communications by Satellite. Prentice-Hall, 1977.
12. Pratt, W.K.: Laser Communication Systems. Wiley, 1969.
13. Siegman, A.E.: The Antenna Properties of Optical Heterodyne Receivers. Proc. IEEE, vol. 54, no. 10, Oct. 1966, pp. 1350-1356.
14. Held, K.J.; and Barry, J.D.: Precision Optical Pointing and Tracking From Spacecraft With Vibrational Noise. Optical Technologies for Communication Satellite Applications, K. Bhasin, ed., Proc. SPIE, Vol. 616, 1986, pp. 160-173.
15. Ponchak, D., et al.: A Technology Assessment of Alternative Communication Systems for the Space Exploration Initiative. NASA TM-103243, 1990.
16. Nelson, R.D.; Ebben, T.H.; Marshalek, R.G.: Experimental Verification of the Pointing Error Distribution of an Optical Intersatellite Link. Free-Space Laser Communication Technologies, G.A. Koepfand and D.L. Begley, eds., Proc. SPIE Vol. 885, pp. 132-142.
17. Lasers and Optronics Buying Guide, Gordon Publications, Dover, DE, 1990.
18. Laser Focus World, vol. 26, no. 5, May 1990, p. 34.
19. Laser Focus World, vol. 26, no. 7, July 1990, p. 11.
20. Aviation Week and Space Technology, vol. 132, no. 23, June 4, 1990.
21. Ross, M.: History of Space Laser Communications Free-Space Laser Communication Technologies, G.A. Koepf and D.L. Begley, eds., Proc. SPIE, Vol. 885, 1988, pp. 2-9.
22. Reiland, W.; Englisch, W.; and Endemann, M.: Optical Intersatellite Communication Links: State of CO_2 Laser Technology. Optical Technologies for Communication Satellite Applications, K. Bhasin, ed., Proc. SPIE, Vol. 616, 1986, pp. 69-76.
23. Reiland, W.; Popescu, A.: On-ground Experiment of a Full-Scale CO_2 Laser Transceiver for Free-Space Communications. Free-Space Laser Communication Technologies, G.A. Koepf and D.L. Begley, eds., Proc. SPIE, Vol. 885, pp. 164-169.
24. Bruno, R.; and Weinberg, A.: Microwave vs Optical ISL Study. Stanford Telecommunications, Inc., Reston, VA, performed under NASA Contract NAS 3-25091, Sept., 1990.
25. Nurre, G.S.; and Dougherty, H.J.: The Pointing System for Space Telescope. National Symposium and Workshop Optical Platforms, C.L. Wyman, ed., Proc. SPIE, Vol. 493, 1984, pp. 22-31.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	January 1993	Technical Paper

4. TITLE AND SUBTITLE

Optical Communications and a Comparison of Optical Technologies for a High Data Rate Return Link From Mars

5. FUNDING NUMBERS

WU-316-30-19

6. AUTHOR(S)

Rodney L. Spence

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Lewis Research Center
Cleveland, Ohio 44135-3191

8. PERFORMING ORGANIZATION REPORT NUMBER

E-6030

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, D.C. 20546-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

NASA TP-3180

11. SUPPLEMENTARY NOTES

Responsible person, Rodney L. Spence, (216) 433-3464.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified - Unlimited
Subject Category 17

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

The important principles of direct- and heterodyne-detection optical free-space communications are reviewed. Signal-to-noise-ratio (SNR) and bit-error-rate (BER) expressions are derived for both the direct-detection and heterodyne-detection optical receivers. For the heterodyne system, performance degradation resulting from received-signal and local-oscillator-beam misalignment and laser phase noise is analyzed. Determination of interfering background power from local- and extended-background sources is discussed. The BER performance of direct- and heterodyne-detection optical links in the presence of Rayleigh-distributed random pointing and tracking errors is described. Finally, several optical systems employing Nd:YAG, GaAs, and CO_2 laser sources are evaluated and compared to assess their feasibility in providing high-data-rate (10- to 1000-Mbps) Mars-to-Earth communications. It is shown that the root mean square (rms) pointing and tracking accuracy is a critical factor in defining the system transmitting laser-power requirements and telescope size and that, for a given rms error, there is an optimum telescope aperture size that minimizes the required power. The results of the analysis conducted indicate that, barring the achievement of extremely small rms pointing and tracking errors (<0.2 μrad), the two most promising types of optical systems are those that use an Nd:YAG laser (λ=1.064 μm) and high-order pulse position modulator (PPM) and direct detection, and those that use a CO_2 laser (λ=10.6 μm) and (phase shift keying) (PSK) homodyne modulation and coherent detection. For example, for a PPM order of M=64 and an rms pointing accuracy of 0.4 μrad, an Nd:YAG system can be used to implement a 100-Mbps Mars link with a 40-cm transmitting telescope, a 20-W laser, and a 10-m receiving photon bucket. Under the same conditions, a CO_2 system would require 3-m transmitting and receiving telescopes and a 32-W laser to implement such a link. Other types of optical systems, such as semiconductor laser systems, are impractical in the presence of large rms pointing errors because of the high power requirements of the 100-Mbps Mars link, even when optimal-size telescopes are used.

14. SUBJECT TERMS

Optical communications; Direct detection; Heterodyne detection; Pointing errors; Mars communication

15. NUMBER OF PAGES

64

16. PRICE CODE

A04

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

NASA-Langley, 1993